

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:29:55 ; Search time 41 seconds
(without alignments)
1686.758 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELASAGVRGAA.....VFLPKISKRQRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2731	100.0	519	20	AAV41114
2	2504.5	91.7	437	20	AAV41116
3	2317.5	84.9	537	20	AAV41127
4	2262.5	82.8	522	20	AAV41124
5	2255.5	82.6	522	20	AAV41123
6	1604.5	58.8	667	20	AAV41118
7	1592	58.3	563	20	AAV41125
8	1532	56.1	605	22	AAV41125
9	1523	55.8	594	22	AAV41125
10	1193.5	43.7	450	22	AAV41125

11	1144	41.9	533	20	AAV41128	Ricinus communis s
12	1134	41.5	494	20	AAV41119	Soybean sucrose tr
13	1126	41.2	501	20	AAV41126	Daucus carota sucro
14	1123.5	41.1	521	23	AAE15581	Alternative versio
15	1122	41.1	510	22	AAV41129	A. thaliana SUT4 p
16	1117	40.9	510	23	ABV90930	Herbicidally activ
17	1110	40.6	523	20	AAV41129	Vicia faba sucrose
18	1107	40.5	491	23	ABV93369	Herbicidally activ
19	1106.5	40.5	500	22	AAV41127	L. esculentum SUT4
20	1106.5	40.5	500	22	AAV41127	S. tuberosum SUT4
21	1106.5	40.5	500	22	AAV41127	Herbicidally activ
22	1106.5	40.5	500	22	AAV41127	Herbicidally activ
23	1106.5	40.5	500	22	AAV41127	Herbicidally activ
24	1088	39.8	507	21	AAV41127	Arabidopsis thalia
25	1088	39.8	507	21	AAV41127	Arabidopsis thalia
26	1088	39.8	507	21	AAV41127	Arabidopsis thalia
27	1088	39.8	507	21	AAV41127	Arabidopsis thalia
28	1088	39.8	507	21	AAV41127	Arabidopsis thalia
29	1075	39.4	520	23	AAE15307	S. tuberosum SUT1
30	1070	39.2	512	23	ABV91556	Glycine max sucros
31	1014	37.1	232	20	AAV41127	Herbicidally activ
32	848.5	31.1	400	20	AAV41127	Wheat sucrose tran
33	677	24.8	322	20	AAV41127	Rice sucrose trans
34	653.5	23.9	350	21	AAV41127	Soybean sucrose tr
35	634.5	23.2	326	21	AAV41127	Arabidopsis thalia
36	564.5	20.7	302	21	AAV41127	Arabidopsis thalia
37	393.5	14.4	599	22	ABV60709	Arabidopsis thalia
38	369	13.5	214	21	AAV41127	Drosophila melanog
39	351	12.9	133	20	AAV41127	Arabidopsis thalia
40	350	12.8	190	20	AAV41127	Corn sucrose trans
41	333	12.2	748	22	AAV41127	Arabidopsis thalia
42	303.5	11.1	553	19	AAV1869	Human polyptide
43	303.5	11.1	553	19	AAV1869	Amino acid encoded
44	303.5	11.1	553	19	AAV1869	Prostate tumour sp
45	303.5	11.1	553	21	AAV28527	Protein encoded by
46	303.5	11.1	553	21	AAV28527	Human immunogenic
47	303.5	11.1	553	22	AAU69763	Human prostate CDN
48	303.5	11.1	553	22	AAU69763	Human prostate tum
49	303.5	11.1	553	22	AAU69763	Human prostate-spe
50	303.5	11.1	553	22	AAU69763	Human prostate-spe
51	303.5	11.1	553	22	AAU69763	Human P501S invent
52	303.5	11.1	553	22	AAU69763	Prostate tumour an
53	303.5	11.1	553	23	ABV77575	Human mast cell re
54	303.5	11.1	553	23	ABV77575	Prostate cancer-as
55	303.5	11.1	553	23	ABV77575	Human L1-12 protei
56	303.5	11.1	553	23	ABV77575	Human PROST 03. H
57	303.5	11.1	553	23	ABV77575	Human breast tumou
58	303.5	11.1	553	23	ABV77575	Prostate tumour an
59	293.5	10.7	595	22	AAU61318	Alpha prepro-P501S
60	282	10.3	166	20	AAV41121	Thioredoxin-ubiqui
61	280	10.3	166	20	AAV41121	Vernonia sucrose t
62	246	9.0	429	21	AAV74548	Arabidopsis thalia
63	246	9.0	429	21	AAV74548	Neisseria meningit
64	245	9.0	429	21	AAV74548	Neisseria meningit
65	245	9.0	429	21	AAV74548	Neisseria meningit
66	234.5	8.6	428	21	AAV74546	Neisseria gonorrh
67	227.5	8.3	530	23	AAV50662	Thioredoxin-ubiqui
68	217.5	8.0	371	22	AAU69875	Human prostate CDN
69	217.5	8.0	371	22	AAU69875	P553S splice varia
70	217.5	8.0	371	23	ABV95335	Human P553S splice
71	215.5	7.9	371	22	AAE01362	Human gene 11 enco
72	215.5	7.9	371	23	ABV64105	Human albumin fusi
73	195	7.1	77	23	ABV91473	Herbicidally activ
74	191.5	7.0	400	22	AAU69907	Human prostate pro
75	191.5	7.0	400	22	AAU69907	Rail2-P501S-E2 cons
76	191.5	7.0	400	23	ABV95367	Rail2-P501S-E2 cons
77	177.5	6.5	326	23	ABV77571	Human mast cell re
78	163.5	6.0	619	21	AAV40554	Human ORFX ORF318
79	162.5	6.0	430	22	AAV96747	Putative P. abyss
80	155.5	5.7	469	21	AAV58289	Lung cancer associ
81	153	5.6	70	22	AAV58289	Sucrose transport
82	150	5.5	430	22	AAV58289	Human secreted pro
83	147.5	5.4	490	23	AAE22906	Human transporter

CC The invention provides nucleic acid sequences (AA23124-223135) encoding
CC sucrose transport proteins (AA41114-41125) derived from corn, rice,
CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can
CC be recombinantly expressed by standard recombinant methodology. The
CC invention facilitates studies on carbohydrate metabolism and function in
CC plants, provides genetic tools for the manipulation of these
CC biosynthetic pathways, and provides a means to control carbohydrate
CC transport and distribution in plant cells.

XX SQ Sequence 497 AA;

Query Match 91.7%; Score 2504.5; DB 20; Length 497;
Best Local Similarity 94.2%; Pred. No. 8.2e-249;
Matches 467; Conservative 18; Mismatches 10; Indels 1; Gaps 1;

QY 24 DHVAPISLGRLLAGWAGGVOYQWALQSLTTPYVOTLGLSHALTSMWLCGPIAGLVV 83
DB 3 DHVAPISLGRLLAGWAGGVOYQWALQSLTTPYVOTLGLSHALTSMWLCGPIAGLVV 62
QY 84 QPLVGLYSDRCTSRWRRRPPILGCMILCVAVIVGFSDDIGAAAGDTRKHCSTLYHGPR 143
DB 63 QPLVGLYSDRCTSRWRRRPPILGCMILCVAVIVGFSDDIGAAAGDTRKHCSTLYHGPR 122
QY 144 WHAAIVVVLGFWLLDFSNNTVOGPARAMMADLCHDHGSPSAANSIFCSWMLGNLTLYSSG 203
DB 123 WHAAIVVVLGFWLLDFSNNTVOGPARAMMADLCHDHGSPSAANSIFCSWMLGNLTLYSSG 182
QY 204 STNNHWHKWFPPKTSACEACANLKGAFVAVVFLVCLTTLTIFAKEVPYRANENLPTT 263
DB 183 STNNHWHKWFPPKTSACEACANLKGAFVAVVFLVCLTTLTIFAKEVPYRANENLPTT 241
QY 264 KAGGEVETPTGPLAVLKGKDLPPGMPSVLLTVAITWLSWFFPILYDMDMGREIYHGD 323
DB 242 KANGEVETPTGPLAVLKGKDLPPGMPSVLLTVAITWLSWFFPILYDMDMGREIYHGD 301
QY 324 PKGSNAQISAFNEGVRGAGGLLNSVLGFSFSLIEPCKRKVGPVVVWTSNFMVCVAM 383
DB 302 PKGSNAQISAFNEGVRGAGGLLNSVLGFSFSLIEPCKRKVGPVVVWTSNFMVCVAM 361
QY 384 AATALISFWSLRDYGHVQDAITANASIKAVCLVFLFAFLGVPPLAILYSVPFAVTAQLAAT 443
DB 362 AATALISFWSLRDYGHVQDAITANASIKAVCLVFLFAFLGVPPLAILYSVPFAVTAQLAAT 421
QY 444 RGGGGGLCTGVNLISIVIPQVIAAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLL 503
DB 422 RGGGGGLCTGVNLISIVIPQVIAAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLL 481
QY 504 PKISKROFRAVSAGGH 519
DB 482 PKISKROFRAVSAGGH 497

RESULT 3
AA41127 standard; protein: 537 AA.

XX AA41127;

XX 17-JAN-2000 (first entry)

DE Oryza sativa sucrose transport protein.

XX Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
KW carbohydrate transport; plant cell.

XX Oryza sativa.

XX WO9953068-A2.

XX 21-OCT-1999.

XX 07-APR-1999; 99WO-US07562.

XX 09-APR-1998; 98US-00811148.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX PI Allen SM, Hitz WD, Rafalski JA;
XX WPI: 1999-620432/53.
XX PT New sucrose transport proteins from plants, useful for controlling
PT carbohydrate transport and distribution in plant cells -
XX Example 3; Page 59-61; 64pp; English.

CC The invention provides nucleic acid sequences (AA23124-223135) encoding
CC sucrose transport proteins (AA41114-41125) derived from corn, rice,
CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can
CC be recombinantly expressed by standard recombinant methodology. The
CC invention facilitates studies on carbohydrate metabolism and function in
CC plants, provides genetic tools for the manipulation of these
CC biosynthetic pathways, and provides a means to control carbohydrate
CC transport and distribution in plant cells.

XX SQ Sequence 537 AA;

Query Match 84.9%; Score 2317.5; DB 20; Length 537;
Best Local Similarity 81.2%; Pred. No. 1.6e-229;
Matches 435; Conservative 39; Mismatches 43; Indels 19; Gaps 3;

QY 1 MARGD-----GGQLAELSAGVRG-----AAVVDHVPISLGRLLIAGWVAG 42
DB 1 MARGSGAGGGGGGGGGLLELSVGVGGGARGGGGGEAAAVETAAPISLGRLLISGMVAG 60
QY 43 GVOYGMALQSLTTPYVOTLGLSHALTSMWLCGPIAGLVVQPLVGLYSDRCTSRWRRR 102
DB 61 GVOYGMALQSLTTPYVOTLGLSHALTSMWLCGPIAGLVVQPLVGLYSDRCTSRWRRR 120
QY 103 PFILTCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPWMAAIVVVLGFWLLDFSN 162
DB 121 PYILTCVCLAVVIVGFSADIGYMGDTKEDCSVYHGRMAAIVVVLGFWLLDFSN 180
QY 163 TVQGPARAMMADLCHDHGSPSAANSIFCSWMLGNLTLYSSGSTNNHWHKWFPLKTSACCE 222
DB 181 TVQGPARAMMADLCHDHGSPSAANSIFCSWMLGNLTLYSSGSTNNHWHKWFPLKTSACCE 240
QY 223 ACANLKGAFVAVVFLVCLTTLTIFAKEVPYRANENLPTTKAGEVETPTGPLAVLKG 282
DB 241 ACANLKGAFVAVVFLVCLTTLTIFAKEVPYRANENLPTTKAGEVETPTGPLAVLKG 299
QY 283 FKDLPPGMPSVLLTVAITWLSWFFPILYDMDMGREIYHGDPKGSNAQISAFNEGVRVGA 342
DB 300 FRNLPTGMPSVLIVTGLTWSWFFPILYDMDMGREIYHGDPKGDPQIEAFNQCVRAGA 359
QY 343 FGLLNSVLGFSFSLIEPCKRKVGPVVVWTSNFMVCVAMAATLISFWSLRDYGHVQ 402
DB 360 FGLLNSVLGFSFSLIEPCKRKVGPVVVWTSNFMVCVAMAATLISFWSLRDYGHVQ 419
QY 403 DAITANASIKAVCLVFLFAFLGVPPLAILYSVPFAVTAQLAATRGGGGLCTGVNLISIVIP 462
DB 420 KAITADKSIKAVCLVFLFAFLGVPPLAILYSVPFAVTAQLAATRGGGGLCTGVNLISIVIP 479
QY 463 QVITAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLLPKISKROFRAVSAGG 518
DB 480 QVITAGAGPMDALFEGKNIPAFGVSFAFALVGGVGVFLLPKISKROFRAVSAGG 535

RESULT 4
AA41124 standard; protein: 522 AA.

XX AA41124;

XX 17-JAN-2000 (first entry)

```

XX DE Wheat sucrose transport protein (clone wlm24.pk0015.g11).
XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
XX KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
XX KW carbohydrate transport; plant cell.
XX OS Triticum aestivum.
XX PN W09953068-A2.
XX PD 21-OCT-1999.
XX PF 07-APR-1999; 99WO-US07562.
XX PR 09-APR-1998; 98US-0081148.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PI Allen SM, Hitz WD, Rafalski JA;
XX DR WPI; 1999-620432/53.
XX DR N-PSDB; AA223134.
XX PT New sucrose transport proteins from plants, useful for controlling
XX PT carbohydrate transport and distribution in plant cells -
XX PS Claim 5; Page 54-55; 64pp; English.
XX CC The invention provides nucleic acid sequences (AA223124-223135) encoding
XX CC sucrose transport proteins (AA41114-Y41125) derived from corn, rice,
XX CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can
XX CC be recombinantly expressed by standard recombinant methodology. The
XX CC invention facilitates studies on carbohydrate metabolism and function in
XX CC plants, provides genetic tools for the manipulation of these
XX CC biosynthetic pathways, and provides a means to control carbohydrate
XX CC transport and distribution in plant cells.
XX SQ Sequence 522 AA;

Query Match 82.8%; Score 2262.5; DB 20; Length 522;
Best Local Similarity 80.5%; Pred. No. 7.2e-224;
Matches 424; Conservative 40; Mismatches 50; Indels 13; Gaps 3;

QY 1 MARGGGQGLAELSGVRGAAA-----VVDHVPISLGRLLILAGMVAGVQYQWALQL 52
DB 1 MARGGGNGEVELSVGGGAGAGADAPVD-----ISLGRLLILAGMVAGVQYQWALQL 56
QY 53 SLLTPVQVOTLGLSHALTSPFMLCGPIAGLVQVPLVGLYSDRCTSRWRRRPPILTSCMLI 112
DB 57 SLLTPVQVOTLGLSHALTSPFMLCGPIAGLVQVPLVGLYSDRCTSRWRRRPPILTSCILI 116
QY 113 CVAVIIVGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAMM 172
DB 117 CVAVVVGFSADIGALGDSKECSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAML 176
QY 173 ADLGDHGGPSAANSIFCSNMALGNILGYSSTNNHWHKFPPLKTSACACANLKGAF 232
DB 177 ADLSAQHGPSAANSIFCSNMALGNILGYSSTNNHWHKFPPLRTRACACANLKGAF 236
QY 233 VAVFLVLCVLTTLTFAKEVYPRANENLPTTKAGGEVETPTGPLAVLKGFKDLPMPGS 292
DB 237 VAVFLVLCVLTTLTFAKEVYPRANENLPTTKAGGEVETPTGPLAVLKGFKDLPMPGS 295
QY 293 VLLVTAITWLSNFPFLYDTHMGREIYHGDPKGSNAQISAFNEGVRVGAFCGLLNSVIL 352
DB 296 VLLVTAITWLSNFPFLYDTHMGREIYHGDPKGTPDEANAPQAGVAGAFGLLNSVYL 355
QY 353 GFSSFLIEPCMKRGVPRVWVTSNEMVCVMAATAALISFWSLRDYHGVQDQAITANASTK 412
DB 356 GFSSFLIEPCMKRGVPRVWVTSNEMVCVMAATAALISFWSLRDYHGVQDQAITANASTK 415
QY 413 AVCLVLFAPLGLVPLAILYSVPFAVTAQAATRGGGGLCTGVLNITSIVIPQVITIALGAGP 472

```

```

DB 416 IVSLALFAFLGLPLAILYSVTFVAVTAQAANRCGGQWLCTGVNLTAIAIPOVITIALGAGP 475.
QY 473 WDALFGKGNIPAFGVASAFALVGGVGVVFLLPKISKROFRAVSGGGH 519
DB 476 WDELFGKGNIPAFGVASAFALVGGVGVVFLLPKISKRLQFRAVSGGGH 522

RESULT 5
AA411123
ID AA411123 standard; protein; 522 AA.
XX AC AA411123;
XX DT 17-JAN-2000 (first entry)
XX DE Wheat sucrose transport protein (clone wlm24.pk0015.g11).
XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
XX KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
XX KW carbohydrate transport; plant cell.
XX OS Triticum aestivum.
XX PN W09953068-A2.
XX PD 21-OCT-1999.
XX PF 07-APR-1999; 99WO-US07562.
XX PR 09-APR-1998; 98US-0081148.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PI Allen SM, Hitz WD, Rafalski JA;
XX DR WPI; 1999-620432/53.
XX DR N-PSDB; AA223133.
XX PT New sucrose transport proteins from plants, useful for controlling
XX PT carbohydrate transport and distribution in plant cells -
XX PS Claim 5; Page 51-53; 64pp; English.
XX CC The invention provides nucleic acid sequences (AA223124-223135) encoding
XX CC sucrose transport proteins (AA41114-Y41125) derived from corn, rice,
XX CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can
XX CC be recombinantly expressed by standard recombinant methodology. The
XX CC invention facilitates studies on carbohydrate metabolism and function in
XX CC plants, provides genetic tools for the manipulation of these
XX CC biosynthetic pathways, and provides a means to control carbohydrate
XX CC transport and distribution in plant cells.
XX SQ Sequence 522 AA;

Query Match 82.6%; Score 2255.5; DB 20; Length 522;
Best Local Similarity 80.3%; Pred. No. 3.8e-223;
Matches 421; Conservative 42; Mismatches 54; Indels 7; Gaps 3;

QY 1 MARGGGQGLAELSGVRG-----AAAVVDHVPISLGRLLILAGMVAGVQYQWALQLSLL 55
DB 1 MARGGGNGEVELSVGGGAGAGADAPVD-----ISLGRLLILAGMVAGVQYQWALQLSLL 60
QY 56 TPYVOTLGLSHALTSPFMLCGPIAGLVQVPLVGLYSDRCTSRWRRRPPILTSCMLICVA 115
DB 61 TPYVOTLGLSHALTSPFMLCGPIAGLVQVPLVGLYSDRCTSRWRRRPPILTSCMLICIA 120
QY 116 VIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAMMADL 175
DB 121 VVVVGFSDIGALGDSKECSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAMMADL 180
QY 176 CDHGGPSAANSIFCSNMALGNILGYSSTNNHWHKFPPLKTSACACANLKGAF 235

```

Db 181 SAOHGPSAANSIFCSWMAIGNILYSSGSTNNHKKWFFPLRTRACCEACANLKGAFVLAV 240
 QY 236 VFLVLCITVTLIFAKEVYRANENLPTTKAGGEVETPTGPLAVLKGFKDLPCGMPVLL 295
 Db 241 LVAFCLVITVIFAKEIPIKAIAPLP-TKGNQVEVEPTGPLAVFKGNLPP-MPSVLL 298
 QY 296 VTATWLSWFFIYLDYDWMGRIYHGDPKGSNAQISAFNEGVRVGAFLGSLNSVILGFS 355
 Db 299 VTGLTWLSWFFIYLDYDWMGRIYHGDPKGTDEANAFQAGVRAGAFGLLNSVILGFS 358
 QY 356 SFLEPCMKRVGPRVWVTSNFMVVCVMAATALISFWSLRDHYGVQDAITANASIKAVC 415
 Db 359 SFLEPCCKRLGPRVWVSSNLFVCLSNAAICIIISWMAATOLHYIQHAIATASKEIKIVS 418
 QY 416 LVLEAFGLVPLAILYSVPFAVTAQLAATRGGGGLCTGVNLISIVIPQVIALGAGPWDA 475
 Db 419 LALFAFGIPLAILYSVPFAVTAQLAATRGGGGLCTGVNLIAIVIPQVIAVAGAPWDE 478
 QY 476 LFGKNIPAFGVSFAFALVGVGVVFLLPKTSKRQRAVSAGGH 519
 Db 479 LFGKNIPAFGVSFAFALVGVGVVFLLPKTSKRQRAVSAGGH 522

RESULT 6
 AAY41118
 ID AAY41118 standard; protein; 667 AA.
 XX
 AC AAY41118;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE
 KW Rice sucrose transport protein (clone rls6.pk0076.e2).
 KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
 KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
 KW carbohydrate transport; plant cell.
 XX
 OS Oryza sativa.
 XX
 PN WO9953068-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US07562.
 XX
 PR 09-APR-1998; 98US-0081148.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Allen SM, Hitz WD, Rafalski JA;
 XX
 DR WPI; 1999-620432/53.
 DR N-PSDB; AA223128.
 XX
 PT New sucrose transport proteins from plants, useful for controlling
 PT carbohydrate transport and distribution in plant cells -
 XX
 PS Claim 5; Page 43-45; 64pp; English.

The invention provides nucleic acid sequences (AA223124-223135) encoding
 sucrose transport proteins (AAY41114-Y41125) derived from corn, rice,
 soybean, Vernonia and wheat tissues. The sucrose transport proteins can
 be recombinantly expressed by standard recombinant methodology. The
 invention facilitates studies on carbohydrate metabolism and function in
 plants, provides genetic tools for the manipulation of these
 biosynthetic pathways, and provides a means to control carbohydrate
 transport and distribution in plant cells.

Sequence 667 AA;

Query Match 58.88; Score 1604.5; DB 20; Length 667;
 Best Local Similarity 55.94; Pred. No. 5.5e-156;
 Matches 299; Conservative 84; Mismatches 101; Indels 51; Gaps 5;

QY 33 RLILAGVAGGVOXGVALQLSLTPYVOTLGLSHALTSMFLCGPIAGLVVQPLVGLYSD 92
 Db 133 KLVLAHVAVAGVQGWALQLSLTPYVOTLGLSHALTSMFLCGPIAGLVVQPLVGLYSD 192
 QY 93 RCTSRWRRRRPFIITGCMILICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIVYL 152
 Db 193 KCRSKYGRRRPFILAGLMICFAVTLIGFSADLGVILGDTTTEHCSLYGSRFRAAIFVL 252
 QY 153 GFLLDSSNTVQGPARAMMADLCHHCPASANSFCSWMAIGNILYSSGSTNNHKKW 212
 Db 253 GFWMLDLANNVTQGPARALLADLSGPDQCSANAFCTMMAVGNVLFSSGASGNHKKW 312
 QY 213 PFLTASACCEACANLKGAFVLAVVFLVLCITVTLIFAKEVYR----- 255
 Db 313 PFLTRACCEACSNLKAFLVAVVFLFCMSVTLYFAEEIPLPTDAORLSDSAPLLNGS 372
 QY 256 -----ANENLPTTKAGGE-VETEPTGPLAY-----LKGFK 284
 Db 373 RDDNNASNEPRNGALPNCHTGDGSNVVPANSNAEDSNRENVEVFNDFGPAVLNLTSMR 432
 QY 285 DLPGMPSVLLVTAITWLSWFFIYLDYDWMGRIYHGDPKGSNAQISAFNEGVRVGA 344
 Db 433 HLPFGMYSVLLVMALTWLSWFFIYLDYDWMGRIYHGDPNUGNUSERKAYDNGVREGAF 492
 QY 345 LLLNSVILGSSFLTEPMCRKRVGPRVWVTSNFMVVCVMAATALISFWSLRDHYGVQDA 404
 Db 493 LLLNSVILGSGFLVDPLCLRLMGLARLVWAINFTVVICMLATAILSWISFDLYSKLHHI 552
 QY 405 ITANASIKAVCLVLFAGVPLAILYSVPFAVTAQLAATRGGGGLCTGVNLISIVIPQV 464
 Db 553 IGANKTVKNSALIVFSLGLPLSITYSVPFSVTABLTAGTGGGGLATGVNLAIWVPOI 612
 QY 465 IIALGAGPMDALFGKNIPAFGVSFAFALVGVGVVFLLPKTSKRQRAVSAGGH 519
 Db 613 VVSLGAGPMDALFGGNNVPALASVFSLGAGVLAVLKLPLD-PNSYR--SAGFH 664

RESULT 7
 AAY41125
 ID AAY41125 standard; protein; 563 AA.
 XX
 AC AAY41125;
 XX
 DT 17-JAN-2000 (first entry)
 XX
 DE Wheat sucrose transport protein (clone wlmk1.pk0002.ell).
 XX
 KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
 KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
 KW carbohydrate transport; plant cell.
 XX
 OS Triticum aestivum.
 XX
 PN WO9953068-A2.
 XX
 PD 21-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US07562.
 XX
 PR 09-APR-1998; 98US-0081148.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Allen SM, Hitz WD, Rafalski JA;
 XX
 DR WPI; 1999-620432/53.
 DR N-PSDB; AA223135.
 XX
 PT New sucrose transport proteins from plants, useful for controlling
 PT carbohydrate transport and distribution in plant cells -
 XX
 PS Claim 5; Page 56-57; 64pp; English.

Db 362 VLVNLTSLRHPAMHSVLIVMALTWLPWPFLEFDTDMGMEVYHGDPKGEADENVAY 421
 Qy 335 NEGVRGAFGLLSVLGSSFLIEPMCKRGVPRVWVTSNFMVCVAMATALISWSL 394
 Db 422 NQGVREGAFGLLSVLGSSFLIEPMCKWIGSRVWVSNFVFCMAGTALISVVS 481
 Qy 395 RYHGYQDAITANASIKAVCLVFLFAGLVPLALYSVPFAVTAQLAATRGSGGGLCTGV 454
 Db 482 SAHTGVQHVIGATKSTQIAALVFLSLLGIPLAVTYSVPFSITAEITADAGGGGLAIGV 541
 Qy 455 LNISIVPQVIALGAGPMDALFKGNIPAFGVASAFALGVGVGVVFLPKISKRRQFRAV 514
 Db 542 LNLAIVLPMVYSLGAGPMDALFGGNIIPAFVLASLAALAGAFIAMLRLPLSS-NFK-- 598
 Qy 515 SAGGH 519
 Db 599 STGFH 603

RESULT 9
 AAG80012
 ID AAG80012 standard; Protein; 594 AA.
 AC AAG80012;
 DT 17-JAN-2002 (first entry)
 DE A. thaliana SUT2 protein.
 KW SUT2; sugar concentration; sugar transport; transgenic plant; herbicide;
 KW flowering time; fruiting time; germination rate; cold tolerance;
 KW drought tolerance; sweetness; branching length; photosynthesis;
 KW pesticide; sugar metabolism.
 OS Arabidopsis thaliana.
 XX WO200173086-A2.
 XX 04-OCT-2001.
 XX 26-FEB-2001; 2001WO-EP02148.
 XX 24-MAR-2000; 2000DE-1014672.
 PR 11-OCT-2000; 2000DE-1050233.
 XX (PROM/) FROMMER W.
 XX Frommer W, Ward JM, Weise A, Barker L, Schulze W, Kuehn C;
 WPI; 2001-611639/70.
 DR N-PSDB; AAI68575.
 XX Modifying sugar flow and concentration in plants, useful e.g. for
 XX increasing content of sugars and oils, by altering activity of a low
 XX affinity, high capacity sugar transporter -
 XX Claim 21b; Page 89-91; 102pp; German.

This invention describes a novel method for modifying the sugar flow
 and/or sugar concentration in plant tissue in which the activity of a
 sugar transporter (I) with high transport capacity but low affinity is
 modified. At least one plant cell is transformed with a vector that
 includes a nucleotide sequence (II) expression of which modifies
 transport activity of (I). The transformed cells are then regenerated.
 The method is used to produce transgenic plants that have increased sugar
 or oil contents in sink organs and harvested materials, altered flowering
 and fruiting times, increased germination rate, better cold and drought
 tolerances, increased sweetness, reduced branching length, and/or
 increased rate of photosynthesis. Nucleic acid sequences that encode (I)
 are also useful for identifying modulators, especially inhibitors of
 sugar transport (potentially useful as herbicides and pesticides) or
 interactors that modify sugar transport, as molecular markers in
 hybridization programs and for modifying the affinity of a protein,

CC especially of the SUC/SUT family, for a substrate, especially sucrose.
 CC Also the central cytoplasmic loop of SUT2 is useful for regulation of
 CC signal transduction, especially in sugar metabolism. The method makes
 CC possible targeted alterations in sugar transport. This sequence
 CC represents the Arabidopsis thaliana SUT2 protein described in the method
 CC of the invention.
 XX
 SQ Sequence 594 AA;
 Query Match 55.8%; Score 1523; DB 22; Length 594;
 Best Local Similarity 53.6%; Pred. No. 1.1e-147;
 Matches 288; Conservative 82; Mismatches 117; Indels 50; Gaps 5;
 Qy 30 SLGLILAGMVGVOYQWALQLSLLPYVOTLGLSHALTSPFMLCGIAGLVQPLVGL 89
 Db 59 SLVTLVLCVAVGQFGWALQLSLLPYIOTLGHAFSSFIWLCGPIITGLVQPPVGI 118
 Qy 90 YSDRCTSRWRRRPFIITGCMILICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIV 149
 Db 119 WSDKCTSKYGRRRPFIIVGSPMISIAVIIIGFSADIGYLLGDSKEHCSTFKGTRAAV 178
 Qy 150 YVLGFWLLDFSNVTVOGPARAMMADLCHDHGSPSANSIFCSWMAIGNILGYSSGSTNNWH 209
 Db 179 FIIGFWLLDLANNTVQGPAPALLADLSGPDQRNTANAVFCLMMAIGNILGFSAGASGKM 238
 Qy 210 KWFPFLKTSACCEACANLKGAFVAVVFLVLCVTLVTLFAKEVPVRANE----- 258
 Db 239 EWFPFLTSRACCAACGNLKAFLAVVFLTICTLVITVFAKEIPTSKNPKTRIODSAPLL 298
 Qy 259 -----NLPTTKAGG-----EVETEPTGPLAY----LKG 282
 Db 299 DDLQSKGLEHSLKNGTANGIKYERVERDTEQFGENSEHEQDETYYVDGPGSVLVNLLTS 358
 Qy 283 FKDLPPGNPVSLLVTAITWLSWFPFIYLDYDMGRIYHGDGPKGSAQISAFNEGVRYGA 342
 Db 359 LRHLPPAMHSLVLIYVMAITWLSWFPFIYLDYDMGRIYHGDGPKGSAQISAFNEGVRYGA 418
 Qy 343 FGLLLNSVILGSSFLIEPMCKRGVPRVWVTSNFMVCVAMATALISFWSLRDYHGVYQ 402
 Db 419 LGLLLNSVILGSSFLIEPMCKRGVPRVWVTSNFMVCVAMATALISFWSLRDYHGVYQ 478
 Qy 403 DAITANASIKAVCLVFLFAGLVPLALYSVPFAVTAQLAATRGSGGGLCTGVNLISVIP 462
 Db 479 YIMRGNETTTRTAAVIVFALLGFLAITVSPFSVTAETADSGGGGLAIGVLNLAIVIP 538
 Qy 463 QVIITAGAGPMDALFGKGNIPAFGVASAFALGVGVGVVFLPKISKROFRAVSAGGH 519
 Db 539 QMIVSLGAGPMDQLFGGGLNLPFVLASVAFAGVIALQRLPTLSS-SFK--STGFH 592

RESULT 10
 AAG80013
 ID AAG80013 standard; Protein; 450 AA.
 AC AAG80013;
 DT 17-JAN-2002 (first entry)
 DE L. esculentum SUT2 protein fragment.
 KW SUT2; sugar concentration; sugar transport; transgenic plant; herbicide;
 KW flowering time; fruiting time; germination rate; cold tolerance;
 KW drought tolerance; sweetness; branching length; photosynthesis;
 KW pesticide; sugar metabolism; tomato.
 XX Lycopersicon esculentum.
 OS
 FH Key Location/Qualifiers
 FT Protein. 1.450
 FT /note= "partial N-terminal fragment"
 XX WO200173086-A2.
 XX

PD 04-OCT-2001.
 XX 26-FEB-2001; 2001WO-EP02148.
 XX 24-MAR-2000; 2000DE-1014672.
 PR 11-OCT-2000; 2000DE-1050233.
 XX (FROM/) FROMMER W.
 XX Frommer W, Ward JM, Weise A, Barker L, Schulze W, Kuehn C;
 XX WPI; 2001-611639/70.
 DR N-PSDB; AA168576.
 XX
 PD Modifying sugar flow and concentration in plants, useful e.g. for
 PT increasing content of sugars and oils, by altering activity of a low
 PT affinity, high capacity sugar transporter
 XX
 PS Claim 21b; Page 91-92; 102pp; German.
 XX
 CC This invention describes a novel method for modifying the sugar flow
 CC and/or sugar concentration in plant tissue in which the activity of a
 CC sugar transporter (I) with high transport capacity but low affinity is
 CC modified. At least one plant cell is transformed with a vector that
 CC includes a nucleotide sequence (II) expression of which modifies
 CC transport activity of (I). The transformed cells are then regenerated.
 CC The method is used to produce transgenic plants that have increased sugar
 CC or oil contents in sink organs and harvested materials, altered flowering
 CC and fruiting times, increased germination rate, better cold and drought
 CC tolerances, increased sweetness, reduced branching length, and/or
 CC increased rate of photosynthesis. Nucleic acid sequences that encode (I)
 CC are also useful for identifying modulators, especially inhibitors of
 CC sugar transport (potentially useful as herbicides and pesticides) or
 CC interactors that modify sugar transport, as molecular markers in
 CC hybridization programs and for modifying the affinity of a protein,
 CC especially of the SUC/SUT family, for a substrate, especially sucrose.
 CC Also the central cytoplasmic loop of SUT2 is useful for regulation of
 CC signal transduction, especially in sugar metabolism. The method makes
 CC possible targeted alterations in sugar transport. This sequence
 CC represents the Lycopersicon esculentum SUT2 protein described in the
 CC method of the invention.
 XX
 XX Sequence 450 AA;
 SQ
 Query Match 43.7%; Score 1193.5; DB 22; Length 450;
 Best Local Similarity 51.7%; Pred. No. 6.5e-114;
 Matches 233; Conservative 59; Mismatches 102; Indels 57; Gaps 5;
 QY 123 SDIGALGDTKEHCYSYHGPRHAAIVYVGLFWLDFSNVTVOGPARAMADLCHDHGFS 182
 DB 1 ADIGYLLGDTKEHCSTFKGTRRAAIVFVVGFWMLDLANTVQGPALLADLSGPDQRN 60
 QY 183 AANSIFCSWMAIGNILGYSGSTNNHKKFPELKTACCACANLKGAFVAVVFLVCL 242
 DB 61 TANAVFCSMWAVGNILGFSAGAGGWHFPFPLTRACCPCGNLKAFLVAVVFLTCLT 120
 QY 243 TVTLIFAKEVP-----RA 256
 DB 121 LVTLIFANEVPLSPKQYKRLSDSAPLLDSDPONTGFDLSOSKRELQSVNSVANNESMGV 180
 QY 257 NENLPTT-----KAGVEVETETGPPLAY--LKGFDLPFGMPSVLLVTAITWLSWFPPT 308
 DB 181 ADNSPKNEORPDKDQGSFADSPGAVLVNLLTSLRHLPPAMHSHVLIVMALTWLPWFPEF 240
 QY 309 LYDTDMGREIYHGDPKGSNAQISAFNEGVRCAGFLLNSVLNLFSSFLIEPMCKVKVP 368
 DB 241 LFPTDMGREVYHGDPKGDEAVNAYNQGVREGAFGLLSNVVLSVSSFLIEPMCKWIGS 300
 QY 369 RVYVWT-SNFMVCMVAMATALISFWSURDYHGVYQVDTANASIKAVCLVLFALGVPLAI 428
 DB 301 RLWVAVSNFIVFCMACTAISVWSISATEGVQHVIGATKSTOIAALVVSLLGIPLAV 360
 QY 429 LYSVPAVTAQLAATRGGGGLCTGVLNLSIVIPQVIAALGAGPMDALFGKGNIPAGVA 488

DB 361 TYSVPFESITAEITADAGGGQGLAIGVNLAIYLPQWVYSLAGPMDALFGGGINPAFVLA 420
 QY 489 SAFALVGGVGVFLLPKISKROFRAVSAGGH 519
 DB 421 SLAALAAGIFAMLRPLNLS--NFX--STGCFH 448
 RESULT 11
 ID AAY41128 standard; protein; 533 AA.
 XX AAY41128;
 XX 17-JAN-2000 (first entry)
 XX Ricinus communis sucrose transport protein.
 KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
 KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
 KW carbohydrate transport; plant cell.
 OS Ricinus communis.
 XX WO9953068-A2.
 PN 21-OCT-1999.
 XX 07-APR-1999; 99WO-US07562.
 PR 09-APR-1998; 98US-0081148.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Allen SM, Hitz WD, Rafalski JA;
 DR WPI; 1999-620432/53.
 PT New sucrose transport proteins from plants, useful for controlling
 PT carbohydrate transport and distribution in plant cells -
 XX Example 3; Page 61-62; 64pp; English.
 CC The invention provides nucleic acid sequences (AA23124-23135) encoding
 CC sucrose transport proteins (AAY4114-Y41125) derived from corn, rice,
 CC soybean, Vernonia and wheat tissues. The sucrose transport proteins can
 CC be recombinantly expressed by standard recombinant methodology. The
 CC invention facilitates studies on carbohydrate metabolism and function in
 CC plants, provides genetic tools for the manipulation of these
 CC biosynthetic pathways, and provides a means to control carbohydrate
 CC transport and distribution in plant cells.
 XX Sequence 533 AA;
 SQ
 Query Match 41.9%; Score 1144; DB 20; Length 533;
 Best Local Similarity 45.0%; Pred. No. 1e-108;
 Matches 226; Conservative 99; Mismatches 151; Indels 26; Gaps 8;
 QY 19 AAVVDHVAIPISLGRLLLAGWAGVOYGVWALQSLTPYVOTLGLSHALTSPMLCGPI 78
 DB 25 AGAAEPNSP--LRKVVWVASTAAGIQFGWALQSLTPYVOLLGPHPTWAAFIWLCGPI 82
 QY 79 AGLVQPLVGLYSDRCTSRWGRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSL 138
 DB 83 SGMVQPIVGYHSDRCTSRFGRRRPFIAAGAAFAVLAFLICYAADLGLHSGD----SL 137
 QY 139 YHGPWRHAAIVVYVGLFWLDFSNVTVOGPARAMADLCHDHGFS-----AANSIFCSWMA 193
 DB 138 DKSPKTRATIAIFVWGFILDDVANMLQGPCRALLADL---SCTSKTKRTANALFSFMA 194
 QY 194 LGNILGYSGSTNNHKKFPELKTACCACANLKGAFVAVVFLVCLTVTLIFAKEVP 253
 DB 195 VGNVLGYAAGATHLYKLFPPFTKTACDVCYCANLKSCFFISIVLLSLTVLALSVKEP 254

```

Qy 254 Y---RANENLPTTKAGGEVETEPTCPL--AVLKGFKDLPPQMPSVLLVLTATITWLSWFFPI 308
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 255 WSPQAVDNAEDDTASQASSAQPMPFGEITLGAFAKLNKRPWILLVLTCLNWTANWFFPL 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 309 LYTDWNGREIYHGDPKGSNAQISAFNEGVRVGAFLGLLNSVILGFSFLIEPWCRCV-G 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 315 LFDTDMGREGVYGGSSGSQLKLYDRGVRAGALGLMLNSVLGFTSLGVEVLARGVGG 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 368 PRVWVTSNEMVCVMAATALI-----SFWSLRDHYGVQDAITANASIKAVCLVLAFL 422
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 375 VKRLWGVNFVLA CLAWTLVLTQAESTRREFATVSGAKVPLPPSGVKGALALFAVM 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 423 GVLAILYSVPFAVTAQLAATRGGQGLCTGVNLISIVIPQVILALGAGPMDALFGKGN 482
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 435 GVPQAIYSIFALASIFNTSGAGQGLSLGVNLISIVIPQVILALGAGPMDALFGGNNL 494
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 483 PAFGVASAFALVGVGVFLLP 504
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 495 PAFVGVAAALASGIFALTMLP 516
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
AY41119
ID AAY41119 standard; protein; 494 AA.
XX
AC AAY41119;
XX
DT 17-JAN-2000 (first entry)
XX
DE Soybean sucrose transport protein (clone sfil.pk0001.g1).
XX
KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
KW carbohydrate transport; plant cell.
XX
OS Glycine max.
XX
PN W09953068-A2.
XX
PD 21-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07562.
XX
PR 09-APR-1998; 98US-0081148.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Hitz WD, Rafalski JA;
XX
WPI; 1999-620432/53.
XX
N-PSDB; AAZ23129.
XX
New sucrose transport proteins from plants, useful for controlling
carbohydrate transport and distribution in plant cells -
Claim 5; Page 45-47; 64pp; English.
XX
The invention provides nucleic acid sequences (AAZ23124-223135) encoding
sucrose transport proteins (AAY41114-Y41125) derived from corn, rice,
soybean, Vernonia and wheat tissues. The sucrose transport proteins can
be recombinantly expressed by standard recombinant methodology. The
invention facilitates studies on carbohydrate metabolism and function in
plants, provides genetic tools for the manipulation of these
biosynthetic pathways, and provides a means to control carbohydrate
transport and distribution in plant cells.
XX
Sequence 494 AA;
XX
Query Match 41.5%; Score 1134; DB 20; Length 494;
Best Local Similarity 46.7%; Pred. No. 9.8e-108;
Matches 234; Conservative 87; Mismatches 146; Indels 34; Gaps 11;

```

```

Qy 28 PISLRLILAGWAGVQYQWALQSLTLPYVQTLGLSHALTSEFMWLCGPIAGLVQPLV 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8 PSLRKMILVSSMAAGIQFGWALQSLTLPYVQTLGVPHANASFIWLCGPIAGLVQPIV 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 88 GLYSDRCTSRGRRRPFILTCMCLICVAVIVVGFSSDGAALGDTKEHCSLYHGPRHAA 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 GYSSDRQCQSRGRRRPFILAGSLAVATAVFLIGYAADIGHAAGD-----NLTKTRPRAV 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 148 IVYVLGFWLLDFSNVTGQPARAMMADLC--DHHGPSAANSITFCSMALGNILCYSSGST 205
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 AIFVIGFWLVDVANNMLQGPCRAFLGDUAAGDEKTKRAANAFFSFFMAVGNILGTAAGSY 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 206 NNHWKFPFLKTSACCEACANLKGAFVAVVFLVLCVLTITLIFAKEVYPYRANENLPTTKA 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 DGLHRLFPFTEACNVFCANLKSCEFAIVLLVLTITLITVLTITVLTITVLTITVLTITV 235
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 266 GGEVE-TEPT-----GPLAVLKGFKDLPPQMPSVLLVLTATITWLSWFFPIYDMDMGRE 318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 236 EKETEAEKTHFSCFCELC--AFKGLKRPMMMLVTAVNWIAWFPYFLFDTDMMGRE 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 319 IYHGDPRKSNQAIISAFNEGVRVGAFLGLLNSVILGFSFLIEPWCRCVGP-RVVWVTSNF 377
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 VYGGD-----VQRAYDSGVHAGSLGLMLNAVVLAVMSLAIEPLRGVVGGIKNLWGIYNI 348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 378 MVCVMAATALISFWSLRDHYGVQDAITANAS--IKAVCLVLAFLVGLVPLAILYSVPFA 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 349 LLAICLGMTVLT--KIAHERLLNPALVGNPSLGKIKVGSVMVFSVLGIPLAITFSVPFA 406
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 436 VTAQLAATRGGQGLCTGVNLISIVIPQVILALGAGPMDALFGKGNIPAFGVASAFALVG 495
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 407 LASTYSSTSGAGQGLSLGVNLIAIVVPMIVSTISGPMWDLFGGNNLPAFVLGAAVAVS 466
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 496 GVGVFLLPKISKR-OPRAVS 515
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 467 AILAVLLLTPTPKKADAVRASS 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
AAY41126
ID AAY41126 standard; protein; 501 AA.
XX
AC AAY41126;
XX
DT 17-JAN-2000 (first entry)
XX
DE Daucus carota sucrose transport protein.
XX
KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
KW carbohydrate metabolism; gene manipulation; biosynthetic pathway;
KW carbohydrate transport; plant cell.
XX
OS Daucus carota.
XX
PN W09953068-A2.
XX
PD 21-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07562.
XX
PR 09-APR-1998; 98US-0081148.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Hitz WD, Rafalski JA;
XX
WPI; 1999-620432/53.
XX
New sucrose transport proteins from plants, useful for controlling
carbohydrate transport and distribution in plant cells -
Example 3; Page 58-59; 64pp; English.
XX
The invention provides nucleic acid sequences (AAZ23124-223135) encoding

```



```
CC useful as herbicides.
XX
XX Sequence 510 AA;
XX
XX Query Match 40.9%; Score 1117; DB 23; Length 510;
XX Best Local Similarity 47.4%; Pred. No. 5.8e-106;
XX Matches 225; Conservative 85; Mismatches 145; Indels 20; Gaps 8;
XX
QY 34 LILAGVAGVGVQGWALQSLTTPYVOTGLSHALTSEFMWLCGPITAGLVVQPLVLYSDR 93
DY 43 LKRVASVAGIQGWALQSLTTPYVOTGLSHALTSEFMWLCGPITAGLVVQPLVLYSDR 102
QY 94 CTSRGRRRPFIITGCMILICVAVIVVGFSSDICAALGDKHCHSLYHGPRHAAIVYVLG 153
DY 103 CTSKYGRRRPFIVAGAVALSISVMVIGHAADIGWAFGDREGKIK----PR--AIVAFVLG 156
QY 154 FWLLDSNNTVQGPAMMADLC--DHHGPSAANSIFCSWMLGNILGYSSGSTNNHWHK 211
DY 157 FWLLDVANNMTQPCRALDLTENDNRTRVANGYFSLFMAVGNVLYGATGSYNGWYKI 216
QY 212 FPLKTSACCEACANLKGAFLVAVVFLVLCITVTLIFAKEVPRANENLPTTKAGEVET 271
DY 217 FFKTVACNVECANLKSIFYIDVFIATITLSVAAHEVPLAS----LASEAHGQTS 272
QY 272 EPTGPLAVLKG-FKDLPPGMPSVLLTAITWLSWFFIYLDTDWGMREIYHGDPKGSNAQ 330
DY 273 TDEAFLEIFGTFRYPFGNWIILLTALTWTIGWFPFIYLDTDWGMREIYHGDPEPNIG 329
QY 331 ISAFNEGVRVAGGLLNSVILGFSFSLIEPCKRKVGRVWVTSNFMVVCVMAATALLIS 390
DY 330 -TSYAGVSMGALGLMLNSVFLGITSVLMKLCRWGAFVNGISNILMAICFLGMIITS 388
QY 391 FWSLRDHYGVODAITANASIKAVCLVFLVLAFLVPLAILYSVPPFAVTAQLAATRGSGOGL 450
DY 389 F--VASHLGYIGHE-QPPASIVFAAVLIETILGIPLAITYSVPYALISIRIESLGLQGL 445
QY 451 CTGVNLISVIVPOTIALGAGPMDALFGKGNIPAFGVASAFALVGVGVFLLPK 505
DY 446 SLGLNLAIVIPQVIVSVSGPMDQLFGGNSPALAVGAATFGIGIVAILALPR 500

RESULT 17
AA411129
ID AAY41129 standard; protein; 523 AA.
XX
XX AC AAY41129;
XX
XX DT 17-JAN-2000 (first entry)
XX
XX DE Vicia faba sucrose transport protein.
XX
XX KW Sucrose transport protein; corn; rice; soybean; Vernonia; wheat;
XX carbohydrate metabolism; gene manipulation; biosynthetic pathway;
XX carbohydrate transport; plant cell.
XX
XX OS Vicia faba.
XX
XX PN W09953068-A2.
XX
XX PD 21-OCT-1999.
XX
XX PF 07-APR-1999; 99WO-US07562.
XX
XX PR 09-APR-1998; 98US-0081148.
XX
XX PA (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX
XX PI Allen SM, Hitz WD, Rafalski JA;
XX
XX DR WPI; 1999-620432/53.
XX
XX PT New sucrose transport proteins from plants, useful for controlling
carbohydrate transport and distribution in plant cells -
```

```
XX
XX PS
XX Example 3; Page 63-64; 64pp; English.
XX
XX The invention provides nucleic acid sequences (AA23124-223135) encoding
XX sucrose transport proteins (AA41114-Y41125) derived from corn, rice,
XX soybean, Vernonia and wheat tissues. The sucrose transport proteins can
XX be recombinantly expressed by standard recombinant methodology. The
XX invention facilitates studies on carbohydrate metabolism and function in
XX plants, provides genetic tools for the manipulation of these
XX biosynthetic pathways, and provides a means to control carbohydrate
XX transport and distribution in plant cells.
XX
XX Sequence 523 AA;
XX
XX Query Match 40.6%; Score 1110; DB 20; Length 523;
XX Best Local Similarity 44.2%; Pred. No. 3.1e-105;
XX Matches 221; Conservative 104; Mismatches 155; Indels 20; Gaps 8;
XX
QY 28 PISLGRLLIAGVAGVQGWALQSLTTPYVOTGLSHALTSEFMWLCGPITAGLVVQPLV 87
DY 31 PSLPKIMVVASIAAGVQGWALQSLTTPYVOTGLSHALTSEFMWLCGPITAGLVVQPLV 90
QY 88 GLYSRDCRSRGRRRPFIITGCMILICVAVIVVGFSSDICAALGDKHCHSLYHGPRHAA 147
DY 91 GYHSRDCRSRGRRRPFIITGCMILICVAVIVVGFSSDICAALGDKHCHSLYHGPRHAA 145
QY 148 IYVVLGFWLLDSNNTVQGPAMMADLC--DHHGPSAANSIFCSWMLGNILGYSSGST 205
DY 146 GFIVVGVFWLLDVANNMTQPCRALDLGDCAGNQRTRNANAFSFMVGNVLYGAGAY 205
QY 206 NNHMKWFPLKTSACCEACANLKGAFLVAVVFLVLCITVTLIFAKEVPRANENLPTTKA 265
DY 206 SKLYHVFPTTKACNVYCANLKSCFFLSIALTLVLTSLATLYVKTALTPKTVVTTED 265
QY 266 GGEVETEPT-GPLAVLKGFKDLPPGMPSVLLTAITWLSWFFIYLDTDWGMREIYHGD 324
DY 266 GSGSGMPCFGOLS--GAFKELKRPMMILLVLTCLNIAFWFFLLDFTDWMGKEV---- 319
QY 325 KGSNAQISAFNEGVRVAGGLLNSVILGFSFSLIEPCKRKV-GPRVWVTSNFMVVCVAM 383
DY 320 GGTGEGHAYDMGVREGALGLMLNSVILGATSLGVDILARGVGVKRLWGIYNFLLAICL 379
QY 384 AATALISFWS--LRDY---HGVQDQAITANASIKAVCLVFLVLAFLVPLAILYSVPPFA 438
DY 380 GLTVLVTKLAAQHSROYAPGTGALGDPPLPSEGIKAGALTFLSVLGVPLAITYSIPFALAS 439
QY 439 QLAATRGSGOGLCTGVNLISVIVPOTIALGAGPMDALFGKGNIPAFGVASAFALVGVV 498
DY 440 IFSSTSGAGOGSLGLVNLAIVIPQVIVSVSGPMDALFGGNSPALAVGAATFGIGIVAIL 499
QY 499 GVFLLPKISKROFRAVSAGG 518
DY 500 SIILLPSPPPDMAKSVSATG 519

RESULT 18
ABB93369
ID ABB93369 standard; Protein; 491 AA.
XX
XX AC ABB93369;
XX
XX DT 31-MAY-2002 (first entry)
XX
XX DE Herbicidally active polypeptide SEQ ID NO 2580.
XX
XX KW Herbicidal; plant; agriculture; herbicide.
XX
XX OS Arabidopsis thaliana.
XX
XX PN WO200210210-A2.
XX
XX PD 07-FEB-2002.
XX
```


Db 130 --IKVRAIAAFVGVFWLLDVANNMTQGPCRALLADLTOKDHRRTVRVANAFLSLEMAIGNI 187
 Qy 198 LGYSSGNNHKKWPFELKTSACCACANLKGAFVAVVFLVCLTTLIFAKEVPYRAN 257
 Db 188 LGFATGSYSGWFKIPFPTLNTACTINCANLKAFAFIIDIFIATTCISISANEQPLDPS 247
 Qy 258 ENLPPTK--AGGEVETPTGTLAVLKG--FKDLPFGMPVSVLLVTAITWLSWFPFILIYDWD 314
 Db 248 RGSSHTREEIGESSHGOEAEFLWELFGIFKYFPGVWVWILLVTAITWIGWFPFLFDWD 307
 Qy 315 MGREIYHGDPK--GSAQAISAFNEGVRGAFGLLLNSVILGFSSFLIEPMCKRKGVRVWV 373
 Db 308 FGREIYGEPPNDGKN-----YSAGVRMGSLGLMLNSVLLGLTSLFMEKLCRKWGAGFTWG 362
 Qy 374 TSNMVVCVMAATALISFWSLRDHYGVQDAITANA-----SIKAVCLVLFPAFLGV 424
 Db 363 VSNVMSLCFTAMLIIT-----AVRSNIDIGQGLPPDGIVIAALVFSILGI 409
 Qy 425 PLAILYSVPFAVTAQAATRGGGGLCTGVNLNISTIVIPQVIAIALGAGPMDALFGKGNIPA 484
 Db 410 PLAITYSVPYALVSSRIDALGLGQGLSMGLNLALVFPQIVVSLGSGPWFDELFGGNSPA 469
 Qy 485 FGVASAFALVGVGVFLLPK 505
 Db 470 FVVAALSAFAGGLIAILAIPR 490

RESULT 20

AAG80017 standard; Protein: 500 AA.

XX AAG80017;

XX 17-JAN-2002 (first entry)

XX S. tuberosum SUT4 protein.

XX SUR4; sugar concentration; sugar transport; transgenic plant; herbicide;
 XX flowering time; fruiting time; germination rate; cold tolerance;
 XX drought tolerance; sweetness; branching length; photosynthesis;
 XX pesticide; sugar metabolism; potato.

XX Solanum tuberosum.

XX WO200173086-A2.

XX 04-OCT-2001.

XX 26-FEB-2001; 2001WO-EP02148.

XX 24-MAR-2000; 2000DE-1014672.

XX 11-OCT-2000; 2000DE-1050233.

XX (FROM/) FROMMER W.

XX Frommer W, Ward JM, Weise A, Barker L, Schulze W, Kuehn C;

XX WPI; 2001-611639/70.

XX N-PSDB; AAI68592.

XX Modifying sugar flow and concentration in plants, useful e.g. for
 XX increasing content of sugars and oils, by altering activity of a low
 XX affinity, high capacity sugar transporter

XX Claim 19b; Page 98-99; 102pp; German.

XX This invention describes a novel method for modifying the sugar flow
 XX and/or sugar concentration in plant tissue in which the activity of a
 XX sugar transporter (I) with high transport capacity but low affinity is
 XX modified. At least one plant cell is transformed with a vector that
 XX includes a nucleotide sequence (II) expression of which modifies
 XX transport activity of (I). The transformed cells are then regenerated.

CC The method is used to produce transgenic plants that have increased sugar
 CC or oil contents in sink organs and harvested materials, altered flowering
 CC and fruiting times, increased germination rate, better cold and drought
 CC tolerances, increased sweetness, reduced branching length, and/or
 CC increased rate of photosynthesis. Nucleic acid sequences that encode (I)
 CC are also useful for identifying modulators, especially inhibitors of
 CC sugar transport (potentially useful as herbicides and pesticides) or
 CC interactors that modify sugar transport, as molecular markers in
 CC hybridization programs and for modifying the affinity of a protein,
 CC especially of the SUC/SUT family, for a substrate, especially sucrose.
 CC Also the central cytoplasmic loop of SUT2 is useful for regulation of
 CC signal transduction, especially in sugar metabolism. The method makes
 CC possible targeted alterations in sugar transport. This sequence
 CC represents the Solanum tuberosum SUT4 protein described in the method of
 CC the invention.

XX Sequence 500 AA;

Query Match 40.5%; Score 1106.5; DB 22; Length 500;

Best Local Similarity 44.9%; Pred. No. 6.8e-105;

Matches 225; Conservative 86; Mismatches 149; Indels 41; Gaps 9;

Qy 21 AVVDHVAP-ISGLRLLAGWAGGVQYCWALQSLTLTPYVQTLGSLHALTSFWMWLCGPIA 79

Db 15 AIREPVKPRVPLRLFLFRVASVAGGIFQGWALQSLTLTPYVQELGPHAWASLIWLCGPLS 74

Qy 80 GLVQPLVGLYSDRCTSRWRRRPILTCMLICVAVTVWGVFSSDIGAALGTKEHCSLY 139

Db 75 GLLVQPLVGHMSDKCTSRGRRPPIVAGAVSIMAVLIIGFSDAGLWLLGDGE----- 129

Qy 140 HGPRHAAIVYVLGFWLLDFSNNTVOGPARAMADLC--DHHGPSAANSIFCSWMAIGNI 197

Db 130 --IKVRAIAAFVGVFWLLDVANNMTQGPCRALLADLTOKDHRRTVRVANAFLSLEMAIGNI 187

Qy 198 LGYSSGNNHKKWPFELKTSACCACANLKGAFVAVVFLVCLTTLIFAKEVPYRAN 257

Db 188 LGFATGSYSGWFKIPFPTLNTACTINCANLKAFAFIIDIFIATTCISISANEQPLDPS 247

Qy 258 ENLPPTK--AGGEVETPTGTLAVLKG--FKDLPFGMPVSVLLVTAITWLSWFPFILIYDWD 314

Db 248 RGSSHTREEIGESSHGOEAEFLWELFGIFKYFPGVWVWILLVTAITWIGWFPFLFDWD 307

Qy 315 MGREIYHGDPK--GSAQAISAFNEGVRGAFGLLLNSVILGFSSFLIEPMCKRKGVRVWV 373

Db 308 FGREIYGEPPNDGKN-----YSAGVRMGSLGLMLNSVLLGLTSLFMEKLCRKWGAGFTWG 362

Qy 374 TSNMVVCVMAATALISFWSLRDHYGVQDAITANA-----SIKAVCLVLFPAFLGV 424

Db 363 VSNVMSLCFTAMLIIT-----AVRSNIDIGQGLPPDGIVIAALVFSILGI 409

Qy 425 PLAILYSVPFAVTAQAATRGGGGLCTGVNLNISTIVIPQVIAIALGAGPMDALFGKGNIPA 484

Db 410 PLAITYSVPYALVSSRIDALGLGQGLSMGLNLALVFPQIVVSLGSGPWFDELFGGNSPA 469

Qy 485 FGVASAFALVGVGVFLLPK 505

Db 470 FVVAALSAFAGGLIAILAIPR 490

RESULT 21

ABB91749

ID ABB91749 standard; Protein: 492 AA.

XX ABB91749;

XX 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 960.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146386.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	13-AUG-1999;	99US-0148684.
PR	16-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	04-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 39.8%; Score 1088; DB 21; Length 512;		
Best Local Similarity 43.6%; Pred. No. 5.6e-103;		
Matches 220; Conservative 100; Mismatches 159; Indels 26; Gaps 10;		
Qy	12	LSAGVRGAAVVDHVA-----PISLGRLLAGWAGVGVGNALQSLSLTTPVQTLGSHA 67
Db	6	MEKAANGASALETQTGELDQPERLKRKIISVSSIAAGVGFGNALQSLSLTTPVQTLGIPKH 65
Qy	68	LTSEFWLGGPIAGLVQPLVGLYSDRCTSRWRRRPFILTCMLTCVAVIVVGFSSDIGA 127
Db	66	WASLIWLCGPISGMLVQPIVGYHSDRCTSRFGRRRPFIVAGAGLVTVAVFLIGYAADIGH 125
Qy	128	ALGDTKEHCSLYHGPRHAAIVYVGLFWLLDFSNNTVGVGPARAMMADLCHHGSPS--AAN 185
Db	126	SMGD-----QLDKPKPTRAIAIFALGFGLVDVANNLTGGPCRAFLADISAGNAKKTRTAN 180
Qy	186	SIFCSWMLGNILYSSGSTNNHKKWFPFLKTSACCEACANLKGAFVAVVFLVLCITVT 245
Db	181	AFFEFFMAVGNVLGYAAGSYRNLYKVPFTWTESCDLYCANLKTCFFISITLLILVTFVS 240
Qy	246	LIFAKEVPYRANENLPTTKAGGEVETEPTGPLAVILKGFKDLPPGMPSVLLYTAITWLSWF 305
Db	241	LCYVKEKFW-----TPEPTADGKASNVPFFG-EIFGAFKELKRRPMMLLIIVTALNIAWF 294
Qy	306	PFILYDWDWGREIYHG--DPKGSNAQISAFNEGVRVCAFGLLNSVLGFSFSLIEBPMC 363
Db	295	PFLFLDWDWGREYVGGNSDATATAASKKLYNDGVRAGALGLMLNAIVLGFMSLGVEMIG 354
Qy	364	RKV-GPRVVVWTSNFMVCVANAATALISFWS--LRDYHGVYQDAITANASTKACVCLVLF 419
Db	355	RKLGAKRLWGIIVNFILAIACIAMTVVVTQKAEHNRD-HGGAKTGPPGN--VTGALTLF 411
Qy	420	AFLGVPLAILYVFPFAVTAQLAATRGCGQLCTGVNLNISIVIPQVIIALGAGPMDALFGK 479

Db 412 AILGIPQAITFSIPFALASIFSTNSGAGLSGLVNLAIWVPMVISVGGPFDELEGG 471
QY 480 GNIPAFGVASAFALYGVGVFLLP 504
Db 472 GNIPAFVLGAIAAVSGVIALTVLP 496

Search completed: April 17, 2003, 11:31:22
Job time : 46 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:30:00 : Search time 17 Seconds
(without alignments)
898.265 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGGGQLAELSGVRGAA.....VELLPKISKROFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1079	39.5	516	1	US-08-356-340-4
2	1079	39.5	516	2	US-08-786-555-4
3	1075	39.4	525	1	US-08-356-340-2
4	1075	39.4	525	2	US-08-786-555-2
5	303.5	11.1	553	4	US-09-020-956-113
6	303.5	11.1	553	4	US-09-030-607-113
7	303.5	11.1	553	4	US-09-605-785-113
8	303.5	11.1	553	4	US-09-439-313-113
9	303.5	11.1	553	4	US-09-352-616A-113
10	303.5	11.1	553	4	US-09-602-877A-101
11	303.5	11.1	553	4	US-09-232-149A-113
12	217.5	8.0	371	4	US-09-605-785-708
13	126.5	4.6	618	4	US-08-595-553A-2
14	114	4.2	435	6	5268463-9
15	112	4.1	467	4	US-09-134-001C-3020
16	112	4.1	528	2	US-08-403-852B-21
17	112	4.1	528	3	US-08-510-646B-22
18	112	4.1	528	4	US-09-231-818-21
19	106	3.9	457	2	US-08-882-704A-6
20	106	3.9	457	4	US-09-151-957-6
21	105.5	3.9	456	6	5432081-10
22	103.5	3.8	474	4	US-09-332-041-5
23	101	3.7	563	2	US-09-031-392-2
24	101	3.7	563	4	US-09-299-549-2
25	101	3.7	563	4	US-09-610-417-2
26	100	3.7	323	4	US-09-134-001C-4635
27	100	3.7	457	6	5268463-7

486	3.7	100	28	4	US-09-291-922-10	Sequence 10, Appl
620	3.7	100	29	1	US-08-301-722A-2	Sequence 2, Appl
456	3.6	99.5	30	6	5432081-7	Patent No. 5432081
255	3.6	99	31	4	US-09-071-710-36	Sequence 36, Appl
255	3.6	99	32	4	US-09-525-397-36	Sequence 2, Appl
619	3.6	99	33	1	US-07-762-132A-2	Sequence 4, Appl
619	3.6	99	34	1	US-08-301-722A-4	Sequence 4837, Ap
326	3.6	97	35	4	US-09-134-001C-4837	Sequence 3299, Ap
411	3.6	97	36	6	5268463-8	Patent No. 5268463
426	3.6	97	37	6	5432081-9	Sequence 26, Appl
539	3.6	96	38	4	US-09-291-922-26	Sequence 7, Appl
434	3.5	95.5	39	2	US-08-677-049-7	Sequence 2, Appl
363	3.5	95	40	1	US-08-148-209A-2	Sequence 9, Appl
635	3.5	95	41	1	US-07-879-617A-9	Sequence 9, Appl
635	3.5	95	42	1	US-08-753-985-9	Sequence 3, Appl
353	3.5	94.5	43	6	5432081-9	Patent No. 5432081
428	3.4	93.5	44	4	US-09-134-001C-5457	Sequence 3457, Ap
524	3.4	93.5	45	4	US-09-134-001C-3960	Sequence 3960, Ap
347	3.4	92.5	46	1	US-08-148-209A-3	Sequence 3, Appl
363	3.4	92	47	1	US-08-118-270-46	Sequence 46, Appl
316	3.4	91.5	48	1	PCT-US93-08528-46	Sequence 3907, Ap
437	3.4	91.5	49	5	US-09-134-001C-3907	Sequence 5, Appl
599	3.4	91.5	50	1	US-08-301-722A-5	Sequence 11, Appl
599	3.3	90.5	51	1	US-08-295-814E-11	Sequence 4, Appl
599	3.3	90.5	52	3	US-08-240-783B-4	Sequence 4, Appl
599	3.3	90.5	53	3	US-09-084-813-4	Sequence 11, Appl
599	3.3	90.5	54	4	US-09-343-361-11	Sequence 4, Appl
599	3.3	90.5	55	5	PCT-US92-09662-4	Sequence 4, Appl
616	3.3	90	56	1	US-07-879-617A-13	Sequence 13, Appl
616	3.3	90	57	1	US-08-753-985-13	Sequence 13, Appl
364	3.3	89.5	58	4	US-09-077-675A-16	Sequence 16, Appl
667	3.3	89	59	1	US-07-879-617A-8	Sequence 8, Appl
667	3.3	89	60	1	US-08-753-985-8	Sequence 2, Appl
388	3.2	88.5	61	1	US-08-087-772A-2	Sequence 5, Appl
400	3.2	88.5	62	1	US-08-351-473B-5	Sequence 4, Appl
400	3.2	88.5	63	1	US-08-450-962-4	Sequence 4, Appl
400	3.2	88.5	64	4	US-08-362-512A-4	Sequence 4, Appl
493	3.2	88.5	65	1	US-08-964-939-4	Sequence 4, Appl
493	3.2	88.5	66	4	US-09-134-001C-5101	Sequence 5101, Ap
415	3.2	88	67	4	US-09-030-267-5	Sequence 5, Appl
418	3.2	88	68	4	US-08-996-644-2	Sequence 2, Appl
1308	3.2	88	69	2	US-09-352-552-2	Sequence 2, Appl
1308	3.2	88	70	3	US-08-118-270-76	Sequence 76, Appl
298	3.2	87.5	71	1	PCT-US93-08528-76	Sequence 76, Appl
298	3.2	87.5	72	5	US-09-031-392-7	Sequence 7, Appl
500	3.2	87.5	73	2	US-09-299-549-7	Sequence 7, Appl
500	3.2	87.5	74	4	US-09-610-417-7	Sequence 7, Appl
500	3.2	87.5	75	4	US-08-295-814E-4	Sequence 4, Appl
627	3.2	87.5	76	1	US-08-291-299-10	Sequence 10, Appl
627	3.2	87.5	77	1	US-08-343-361-4	Sequence 4, Appl
627	3.2	87.5	78	5	PCT-US93-01959-4	Sequence 4, Appl
627	3.2	87.5	79	5	PCT-US95-10579-10	Sequence 10, Appl
632	3.2	87.5	80	1	US-08-295-814E-10	Sequence 10, Appl
632	3.2	87.5	81	1	US-09-343-361-10	Sequence 10, Appl
632	3.2	87.5	82	3	PCT-US93-01959-10	Sequence 10, Appl
4551	3.2	85	83	3	US-09-320-878-1	Sequence 31, Appl
4613	3.2	87.5	84	4	US-09-105-537-31	Sequence 6, Appl
11877	3.2	87.5	85	4	US-09-463-238-4	Sequence 4, Appl
87	3.2	87	86	4	US-09-134-001C-3564	Sequence 3487, Ap
87	3.2	86.5	87	4	US-09-031-392-6	Sequence 6, Appl
473	3.2	86.5	88	2	US-09-299-549-6	Sequence 6, Appl
473	3.2	86.5	89	4	US-09-610-417-6	Sequence 6, Appl
553	3.2	86.5	90	3	US-08-501-572-3	Sequence 3, Appl
553	3.2	86.5	91	3	US-09-040-444-3	Sequence 10, Appl
559	3.2	86.5	92	1	US-07-879-617A-10	Sequence 10, Appl
559	3.2	86.5	93	1	US-08-753-985-10	Sequence 12, Appl
614	3.2	86.5	94	1	US-08-295-814E-12	Sequence 4004, Ap
614	3.2	86.5	95	4	US-09-343-361-12	
349	3.1	86	96	4	US-09-134-001C-4004	

ALIGNMENTS

RESULT 1

US-08-356-340-4

; Sequence 4, Application US/08356340

; Patent No. 5608146

; GENERAL INFORMATION:

; APPLICANT: FROMMER, Wolf-Bernd

; APPLICANT: RIESMEIER, Jorg

; TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE

; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA AND PLANTS CONTAINING A

; TITLE OF INVENTION: TRANSPORTER AS WELL AS A PROCESS FOR THE PREPARATION AND

; TITLE OF INVENTION: TRANSFORMATION OF YEAST STRAINS FOR THE IDENTIFI...

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ostroienk, Faber, Gerb & Soffen

; STREET: 1180 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: US

; ZIP: 10036-8403

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/356,340

; FILING DATE: 21-DEC-1994

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US PCT/EP93/01604

; FILING DATE: 22-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: DE P42 20 759,2

; FILING DATE: 24-JUN-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Meilman, Edward A.

; REGISTRATION NUMBER: 24,735

; REFERENCE/DOCKET NUMBER: P/951-106

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 382-0700

; TELEFAX: (212) 382-0888

; TELEX: 236925

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 516 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-356-340-4

Query Match 39.5%; Score 1079; DB 1; Length 516;

Best Local Similarity 43.9%; Pred. No. 6.6e-100;

Matches 225; Conservative 101; Mismatches 162; Indels 24; Gaps 11;

Qy 16 VRGAAVVDHVAPISLGRLLIAGVAGVQYQGWALQSLTLPYVOTLGLSHALTSFMWLC 75

Db 15 VSSSLQVEQPLAPSKLWIIIVASIAAGVQFGWALQSLTLPYVOLLGIPHKFASFIWLC 74

Qy 76 GPIAGLVQPLVGLYSDRCTSRWGRRRRPFILTGCMLICVAVIVVGFSSDGAALGDTKEH 135

Db 75 GPISGMIVQPVVGYSDNCSSRFGRRRPFIAAGAAVIAVFLIGFAADLGHASGDT--- 131

Qy 136 CSLYHGRPRWHAIIIVVLGFWLLDFSNNTVQGPARAMMADLGDHGHPS----AANSIFCSW 191

Db 132 --LGKGFPRATAVFVVGFWILDVANNMLQGPCRALLADLSG--GKSGRMRTANAFSFF 187

Qy 136 CSLYHGRPRWHAIIIVVLGFWLLDFSNNTVQGPARAMMADLGDHGHPS----AANSIFCSW 191

Db 132 --LGKGFPRATAVFVVGFWILDVANNMLQGPCRALLADLSG--GKSGRMRTANAFSFF 187

Qy 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Db 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Qy 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Db 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Qy 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Db 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Qy 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Db 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Db 188 MAVGNILGYAAGSYSHLFKVPFFSKTKACDMYCANLKSCFFIA-IFLLSLTTTALTTLVR 246

Qy 249 AKEVPYRANENLPTTKAG-GEVETPTGPLAVLKGFKDLPPGMPSVLLVTAITWLSWPF 307

Db 247 ENELPEKDEQIDEKLAGAGSKVFFG--EIFGALKELPRPMWILLVTCNLNIAWPF 304

Qy 308 ILYDTDMGRIYHGDPKGSNAQISAFNEGVRVGAFGLLLNSVLGFSFSLIEPMCRKV- 366

Db 305 FLYDTDMWAKEVFGG--QVGDAEL--YDLGVRAGAGMLLLQSVVLGPMSLGVFEFLGKIG 360

Qy 367 GPRVVVTSNPMVCVMAATALISFWSLRD-YHGVQDAITANASIKAVCLVFAFLGVP 425

Db 361 GAKLWGLTNFLVLAICLANTLITVTKMAEKSRQHPAGTGMGTTCVKITGALLFAALGIP 420

Qy 426 LAILVSPFVAVTAOLATRGCGQLCTGVNLISIVIPQVIAIAGAGPMDALFGKGNIPAF 485

Db 421 LAATFSPFALASIFSSNRSGQSLGLVNLAIIVPQMLVSLVGGPMDLFGGGLPGF 480

Qy 486 GVSAFALVGGVGVFLLPKISKROFRAVSAG 517

Db 481 VVGAVAAAAAVALTMLPSPADAKPAVAMG 512

RESULT 2

US-08-786-555-4

; Sequence 4, Application US/08786555B

; Patent No. 5981181

; GENERAL INFORMATION:

; APPLICANT: FROMMER, Wolf-Bernd

; APPLICANT: RIESMEIER, Jorg

; TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE TRANSPORTER,

; TITLE OF INVENTION: PLASMIDS, BACTERIA AND PLANTS CONTAINING A TRANSPORTER

; TITLE OF INVENTION: AS WELL AS A PROCESS FOR PREPARATION AND TRANSFORMATION

; TITLE OF INVENTION: OF YEAST STRAINS FOR THE IDENTIFICATION

; FILE REFERENCE: Frommer

; CURRENT APPLICATION NUMBER: US/08/786,555B

; CURRENT FILING DATE: 1997-01-21

; EARLIER APPLICATION NUMBER: 08/356,340

; EARLIER FILING DATE: 1994-12-21

; EARLIER APPLICATION NUMBER: PCT/EP93/01604

; EARLIER FILING DATE: 1993-06-22

; EARLIER APPLICATION NUMBER: DE P4220759,2

; EARLIER FILING DATE: 1992-06-24

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 516

; TYPE: PRT

; ORGANISM: Spinacia oleracea

; US-08-786-555-4

Query Match 39.5%; Score 1079; DB 2; Length 516;

Best Local Similarity 43.9%; Pred. No. 6.6e-100;

Matches 225; Conservative 101; Mismatches 162; Indels 24; Gaps 11;

Qy 16 VRGAAVVDHVAPISLGRLLIAGVAGVQYQGWALQSLTLPYVOTLGLSHALTSFMWLC 75

Db 15 VSSSLQVEQPLAPSKLWIIIVASIAAGVQFGWALQSLTLPYVOLLGIPHKFASFIWLC 74

Qy 76 GPIAGLVQPLVGLYSDRCTSRWGRRRRPFILTGCMLICVAVIVVGFSSDGAALGDTKEH 135

Db 75 GPISGMIVQPVVGYSDNCSSRFGRRRPFIAAGAAVIAVFLIGFAADLGHASGDT--- 131

Qy 136 CSLYHGRPRWHAIIIVVLGFWLLDFSNNTVQGPARAMMADLGDHGHPS----AANSIFCSW 191

Db 132 --LGKGFPRATAVFVVGFWILDVANNMLQGPCRALLADLSG--GKSGRMRTANAFSFF 187

Qy 192 MALGNILGYSSGSTNNHMKWPFPLKTSACCACANLKGAFVAVVFLVLCIT---VTILIF 248

Db 188 MAVGNILGYAAGSYSHLFKVPFFSKTKACDMYCANLKSCFFIA-IFLLSLTTTALTTLVR 246

Qy 249 AKEVPYRANENLPTTKAG-GEVETPTGPLAVLKGFKDLPPGMPSVLLVTAITWLSWPF 307

Qy	27	APISGLRLIAGWVAGVQYGNALQJLSLLPPYVOTLGLSHALTSFWMLCGPIAGLVVQPL	86
Db	31	AEATLKLGLGVSAVGAQFGWALQJLSLLPPYVQQLGHPHTWAAIYTWLCGPISGMIVQPL	90
Qy	87	VGLYSDRCTSRWRRRPFILTCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRWA	146
Db	91	VGYSDRCTSRFGRRRPFIAAGAAVLAVAVGLIGFADIGAASDDTGNVA---KPR--A	145
Qy	147	AIYVVLGFWLLDFSNNTVQGPARAMADLC--DHHGPSAANSIFCSNMALGNLTGYSSGS	204
Db	146	IAVFVVGFWLIDVANNTLQGPCALLADMAAGSQTKRYANAFSFFEMALGNLTGGYAAGS	205
Qy	205	TNNHWKPFPLKTSACCEACANLKGAFLVAVFVLVLCLTVTLFEAKE----VPYRANEN	259
Db	206	YSRLTYVPFPTKTAACDVCYCANLKSCFFISITLLIVLTILALSVYKERTIDETGEED	265
Qy	260	LPT--TKAGGEVETPTGPLAVLKGFDLPDPMGSVLVLVTAITWLSNPPFLYDPTDWMGR	318
Db	266	LKNRNNSSGCARLPFFGQL--IGALKDLKPLMLILLVLTALNIAWFPFLFLFTDWMGKE	323
Qy	319	IYHGDPRKSNAOISAFNEGVRVGAFLGLNLSVTLGSSFLIEPMCRKV-CPRVVVWTSNF	377
Db	324	VY----GGTVGEKLYDQGVHAGALGLMIINSVVLGYWMSLIEGLRWVGAKRLWIVNI	379
Qy	378	MVCYMAATAALI--SEWSLRDYGHYVODAI--TANASIKAVCLVLFLAFLGVLPLAIVSVP	433
Db	380	ILAVCLAMTVLTKSAEHRFSDSHHIMGSVAVPPPPAGVKGGALAIFAVLIGLPLAITSFIP	439
Qy	434	FANTAOLAATRGSGOGLCTGVNLNLSIVIPOVITALGAGPMDALFGKGNIPAFGVASAFAL	493
Db	440	LALAFSAGSGSGGUGLSLGLVNLAIYVPMQFVSTVSGPMDAMFGGNLPAFVVGVAAT	499
Qy	494	VGVGVGVFLLPKISKQFRAVSAGGH	519
Db	500	ASAVLSFTLLPSPPEAKIGSGMGHH	525

```

RESULT 4
US-08-786-555-2
; Sequence 2, Application US/08786555B
; Patent No. 5981181
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; APPLICANT: RIESMEIER, Jorg
; TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE TRANSPORTER,
; TITLE OF INVENTION: PLASMIDS, BACTERIA AND PLANTS CONTAINING A TRANSPORTER
; TITLE OF INVENTION: AS WELL AS A PROCESS FOR PREPARATION AND TRANSFORMATION
; TITLE OF INVENTION: OF YEAST STRAINS FOR THE IDENTIFICATION
; FILE REFERENCE: Frommer
; CURRENT APPLICATION NUMBER: US/08/786,555B
; CURRENT FILING DATE: 1997-01-21
; EARLIER APPLICATION NUMBER: 08/356,340
; EARLIER FILING DATE: 1994-12-21
; EARLIER APPLICATION NUMBER: PCT/EP93/01604
; EARLIER FILING DATE: 1993-06-22
; EARLIER APPLICATION NUMBER: DE P4220759.2
; EARLIER FILING DATE: 1992-06-24
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 525
; TYPE: PR1
; ORGANISM: Spinacia oleracea
US-08-786-555-2

```

Qy 27 APISLGRLLIAGMWAGGVQYWGALQLSLTTPVQVTGLSHALTSFMMWLCGPIAGLVVQPL 86
| : | : | | |||:||||| ||| | : ||||| : |||||
Db 31 AEATKKKLGIVASAAVGQVFQWALQLSLTPYVOLLGIPTWAAAYTWLCGPISGMIVQPL 90

QY 87 VGLYSDRCTSRGRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGRWHA 146
Db 91 VGLYSDRCTSRGRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGRWHA 145
QY 147 ATIVYVGLFWLDFSNVTVQGPARAMADLC--DHGCPAANSIFCSWALGNILGYSSGS 204
Db 146 IAVFVGVFWLDFSNVTVQGPARAMADLC--DHGCPAANSIFCSWALGNILGYSSGS 205
QY 205 TNNHKKWPELTKSACCEACANLKGAFVAVVFLVCLTITLIFAKE-----VPRANEN 259
Db 206 YSLRTVFPFTKTAADVCANLKSCTFISITLLVILALSVVKKRQITIDEIQEED 265
QY 260 LPT-KTAGEVEETEPTGLPLAVLKGDKLPPGMPSPVLLVTATWLSWFFILYDTDMGRE 318
Db 266 LKNRNNSSCARLPFGQL--IGALKDLKPKMLILLVLTALNWIAPFLLFDTDMCKE 323
QY 319 IYHDPKSNQAISAFNEGVVGAFLGLLNSVILGFSSFLTEPMCRKV-GRPVVWVTSNF 377
Db 324 VY-----GGTVGEGKLYDQGVHAGALGLMINSVYLVGMSLSTEGLMVGGAKRLWGI 379
QY 378 MYCVAMAATALT--SFWSLRDHYGVVQDAI--TANASIKAVCLVLFVGLVPLALYVSP 433
Db 380 ILAVCLAMTVLTKSAEHRDSSHIMGSAVPPPPAGVKGGLAIFAVGLPIALITFISIP 439
QY 434 FAVTAQLAATRGCGGLCTGVNLINISIVIPQVIALGAGPMDALFGKINIPAFGVASAFAL 493
Db 440 LALASIFSASSGCGSLGLVNLALVVPQMEVSVTSGPMDAMFGGNLPAPVVGAVAT 499
QY 494 VGVVGVVFLPKISKRPRAVSAGGH 519
Db 500 ASAVLSFTLLSPSPPEAKTGGSMGGH 525

RESULT 5
US-09-020-956-113
; Sequence 113, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020.956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens
US-09-020-956-113
Query Match 11.18; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.38; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;
QY 33 RLILAGVAGVQVQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVVQPLVGLYSD 92
Db 17 QLLVNLTLTGLEVCLAGITYVPPLLEVGVEEKFMVILGIPVGLVLCVPLLGSSASD 76
QY 93 RCTSRWRRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGRWHAIVVYL 152
Db 77 HWRGRRRRPFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYHGRWHAIVVYL 127
QY 153 GFLLLDPSNNIVQGPARAMADLC--DHGCPAANSIFCSWALGNILGYSSGSTNNWH 209
Db 128 GVLLDFGCGVCFPLEALLSDFLRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTITLIFAKEVYPYRANENLPTTK 264
Db 183 -WDTSAALPYLGTQEEC-----LFG--LLTLIFT-CVAATLLVAEEAALGPTG--PAEG 231
QY 265 AGGEVETEPTGLAVLKGDKL-----PPGMPSPVLLVTATWLSWFFILY 310
Db 232 LSAPSLSPHCCPCARLARLAFRLNGLALLPRLHQLCCRPRTLRLFLVAELCSWMLMTFTLF 291
QY 311 DTDMGRRIYHGDPK--GSNAQISAFNEGVVGAFLGLLNSVILGFSSFLTEPMCRKV 367
Db 292 YTFDFGEGLYGVVPAEPGTEAR-RHYDEGVNMGSLGLFLOCAISLVSFLVMDRLVQRFG 350
QY 368 PRVWVTSNFMVCMAMAATALTISFMSLRDHYGVVQDAITANASIKAVCLVLFVGLVPLA 427
Db 351 TRAVVLSAAFPVPAAGATCL-----SHSVAVVTSAAALGTFTFSALQILPYTLA 400
QY 428 ILX 430
Db 401 SLY 403

RESULT 6
US-09-030-607-113
; Sequence 113, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030.607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 113:

SEQUENCE CHARACTERISTICS:
LENGTH: 553 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-030-607-113

Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;
QY 33 RLILAGVAGGVOYQWALQSLTTPYVOTLGLSHALTSPMMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLVNLITFGLVCLAGITVPPLLLLEVGEVEKFTWLVGIVGLVCVPLIGSASD 76
QY 93 RCTSMGRRRRPFILTCMLICVAVIVGVFSSDIGAALGDTKEHCSLYHGRWHAAIVYVL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127
QY 153 GFWLDFSNNTVOGPARMMADLC---DHHGPSAANSIFCSWMLGNILGYSSGSTNNWH 209
DB 128 GVGLLDFCGQVCFPTPLEALLSDLFDPDH--CRQAYSVYAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----PPELKTSAACCEACANLKGAFVAVVFLVCLTTLIFAKEVPYRANENLPTTK 264
DB 193 -WDTSAALPYLGTQEC-----LFG--LTLFLT-CVAATLLVAEEAALGPT--PAG 231
QY 265 AGGEVETEPTGLAYLKGFKDL-----PPGMPSVLLVTAITWLSWPPFIY 310
DB 232 LSAPSLPHCCPCARLARAFNLGALLPRLHQLCCRMPTLRLFLVAELCSWMLMTFTLF 291
QY 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVRVGAAGLLNSVILGFSFLIEPCKRVG 367
DB 292 YTFDVEGELYGVRAEPGTEAR--RHYDEGVRMGLGLFLOCAISLVSLVMDRLVQRF 350
QY 368 PRVWVTNFMVCMVMAATALISFWSLRDYGIVQDAITANASIKAVCLVLFAGVPLA 427
DB 351 TRAVLASVAAPFVAGATCL-----SHSVAVVTASAALTGTFESALQILPYTLA 400
QY 428 ILY 430
DB 401 SLV 403

RESULT 7
US-09-605-785-113
Sequence 113, Application US/09605785
Patent No. 6321716
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C16
CURRENT APPLICATION NUMBER: US/09/605.785

CURRENT FILING DATE: 2000-06-27
NUMBER OF SEQ ID NOS: 835
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 113
LENGTH: 553
TYPE: PRT
ORGANISM: Homo sapien
US-09-605-785-113

Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;
QY 33 RLILAGVAGGVOYQWALQSLTTPYVOTLGLSHALTSPMMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLVNLITFGLVCLAGITVPPLLLLEVGEVEKFTWLVGIVGLVCVPLIGSASD 76
QY 93 RCTSMGRRRRPFILTCMLICVAVIVGVFSSDIGAALGDTKEHCSLYHGRWHAAIVYVL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127
QY 153 GFWLDFSNNTVOGPARMMADLC---DHHGPSAANSIFCSWMLGNILGYSSGSTNNWH 209
DB 128 GVGLLDFCGQVCFPTPLEALLSDLFDPDH--CRQAYSVYAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----PPELKTSAACCEACANLKGAFVAVVFLVCLTTLIFAKEVPYRANENLPTTK 264
DB 193 -WDTSAALPYLGTQEC-----LFG--LTLFLT-CVAATLLVAEEAALGPT--PAG 231
QY 265 AGGEVETEPTGLAYLKGFKDL-----PPGMPSVLLVTAITWLSWPPFIY 310
DB 232 LSAPSLPHCCPCARLARAFNLGALLPRLHQLCCRMPTLRLFLVAELCSWMLMTFTLF 291
QY 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVRVGAAGLLNSVILGFSFLIEPCKRVG 367
DB 292 YTFDVEGELYGVRAEPGTEAR--RHYDEGVRMGLGLFLOCAISLVSLVMDRLVQRF 350
QY 368 PRVWVTNFMVCMVMAATALISFWSLRDYGIVQDAITANASIKAVCLVLFAGVPLA 427
DB 351 TRAVLASVAAPFVAGATCL-----SHSVAVVTASAALTGTFESALQILPYTLA 400
QY 428 ILY 430
DB 401 SLV 403

RESULT 8
US-09-439-313-113
Sequence 113, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang, Yuqui
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
APPLICANT: Day, Craig
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439.313
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 113
LENGTH: 553
TYPE: PRT
ORGANISM: Homo sapien

US-09-439-313-113

```
Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGMAGGVOYQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLTFFGLEVCIAAGITVPPPLLEVGVEEFMTVMVLGIGPVGLVCVPLGGSASD 76
QY 93 RCTSRWRRRPFILTCGMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIVVYL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127
QY 153 GFWLDFSNNTVOGPARAMADLC---DHHGPSAANSIFCSNMALGNILGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPLEALLSDFLRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTITLIFAKEVPYRANENLPTTK 264
DB 183 -WDTLSALAPYLGTQEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGPT--PAEG 231
QY 265 AGGEVETETGPLAVLKGFKDL-----PCMPSVLLVTAITWLSWEPFFILY 310
DB 232 LSAPLSLPHCCPCRARLARFNLGALLPRLHQLCCRPRTLRLFLVAELCSNMALMTFTLF 291
QY 311 DTDWMGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLEPCKRVG 367
DB 292 YTDVFGEGLYQGVPRAEFGTEAR-RHYDEGVRMGSGLFLQCAISLVSFLVMDRLVQRF 350
QY 368 PRVWVTSNFMVVCVMAATATISFWSLRDHYGYVODATITANASIKAVCLVLFAGVPLA 427
DB 351 TRAVVLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILX 430
DB 401 SLY 403

RESULT 9
US-09-352-616A-113
; Sequence 113, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-352-616A-113

Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGMAGGVOYQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLTFFGLEVCIAAGITVPPPLLEVGVEEFMTVMVLGIGPVGLVCVPLGGSASD 76
QY 93 RCTSRWRRRPFILTCGMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIVVYL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127
```

```
QY 153 GFWLDFSNNTVOGPARAMADLC---DHHGPSAANSIFCSNMALGNILGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPLEALLSDFLRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTITLIFAKEVPYRANENLPTTK 264
DB 183 -WDTLSALAPYLGTQEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGPT--PAEG 231
QY 265 AGGEVETETGPLAVLKGFKDL-----PCMPSVLLVTAITWLSWEPFFILY 310
DB 232 LSAPLSLPHCCPCRARLARFNLGALLPRLHQLCCRPRTLRLFLVAELCSNMALMTFTLF 291
QY 311 DTDWMGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLEPCKRVG 367
DB 292 YTDVFGEGLYQGVPRAEFGTEAR-RHYDEGVRMGSGLFLQCAISLVSFLVMDRLVQRF 350
QY 368 PRVWVTSNFMVVCVMAATATISFWSLRDHYGYVODATITANASIKAVCLVLFAGVPLA 427
DB 351 TRAVVLASVAAPFVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
QY 428 ILX 430
DB 401 SLY 403

RESULT 10
US-09-602-877A-101
; Sequence 101, Application US/09602877A
; Patent No. 6432707
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.446C5
; CURRENT APPLICATION NUMBER: US/09/602,877A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 101
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-602-877A-101
```

```
Query Match 11.1%; Score 303.5; DB 4; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-22;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

QY 33 RLILAGMAGGVOYQWALQSLTTPYVOTLGLSHALTSMFMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLTFFGLEVCIAAGITVPPPLLEVGVEEFMTVMVLGIGPVGLVCVPLGGSASD 76
QY 93 RCTSRWRRRPFILTCGMLICVAVIVVGFSSDGAALGDTKEHCSLYHGPWRHAAIVVYL 152
DB 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLAGL-----CP---DPRPLELALLIL 127
QY 153 GFWLDFSNNTVOGPARAMADLC---DHHGPSAANSIFCSNMALGNILGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPLEALLSDFLRDPDH--CRQAYSVAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCACANLKGAFVAVVFLVCLTITLIFAKEVPYRANENLPTTK 264
DB 183 -WDTLSALAPYLGTQEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGPT--PAEG 231
QY 265 AGGEVETETGPLAVLKGFKDL-----PCMPSVLLVTAITWLSWEPFFILY 310
DB 232 LSAPLSLPHCCPCRARLARFNLGALLPRLHQLCCRPRTLRLFLVAELCSNMALMTFTLF 291
QY 311 DTDWMGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLEPCKRVG 367
```

```

Db      292  YTDFVGGGLYQGVPRAPPGTEAR-RHYDEGVRMGSLGLFLQCAISLVSLVMDRLVQREG 350
Qy      368  PRVVWTSNFMVCMVAMAATALISFWSLROHYGVQDAITANASIKAVCLVLFLFAFLGVPILA 427
Db      351  TRAVYLASVAAFPVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
Qy      428  ILY 430
Db      401  SLY 403

RESULT 11
US-09-232-149A-113
: Sequence 113, Application US/09232149A
: Patent No. 6465611
: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davlin C.
: APPLICANT: Mitcham, Jennifer Lynn
: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
: TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
: FILE REFERENCE: 210121.427C6
: CURRENT APPLICATION NUMBER: US/09/232,149A
: CURRENT FILING DATE: 1999-01-15
: NUMBER OF SEQ ID NOS: 338
: SOFTWARE: FastSEQ for Windows Version 3.0
: SEQ ID NO 113
: LENGTH: 553
: TYPE: PRT
: ORGANISM: Homo sapien
US-09-232-149A-113

```

```

: GENERAL INFORMATION:
: APPLICANT: Xu, Jiangchun
: APPLICANT: Dillon, Davin C.
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Jiang, Yuqi
: APPLICANT: Henderson, Robert A.
: APPLICANT: Kalos, Michael D.
: APPLICANT: Fanger, Gary R.
: APPLICANT: Retter, Marc W.
: APPLICANT: Stolk, John A.
: APPLICANT: Day, Craig H.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: Carter, Darrick
: APPLICANT: Li, Samuel
: APPLICANT: Wang, Aijun
: APPLICANT: Skeiky, Yasir A.W.
: APPLICANT: Hepler, William
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
: TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
: FILE REFERENCE: 210421.427C16
: CURRENT APPLICATION NUMBER: US/09/605,785
: CURRENT FILING DATE: 2000-06-27
: NUMBER OF SEQ ID NOS: 835
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 708
: LENGTH: 371
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-605-785-708

```

```

Query Match          11.1%; Score 303.5; DB 4.; Length 553;
Best Local Similarity 25.3%; Pred. No. 6e-24;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps

QY      33 RLILAGVAGGVGYGWAQLQSLLTPYVOTLGLSHALTSEFMWLCGPITAGLVWPVLGCLYS D 92
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      17 QLLLVNLTFTGLEVCVLAAGITYVPPLLEVGVEEKFTMTVGLGIPVGLVCVPLLGASD 76

QY      93 RCTSRWGRRRPFILTGCMLICVAVIVGFSSDIGAALGDTKHCSLYHGPRHAAIYYVL 152
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      77 HWRGRVGRRRPFIWALSGLTLLSLFLIPRAGNLAGLL-----CP----DPRPLEALLIL 127

QY      153 GFWLIDFSNNTVGGPARAMMADLC---DRHGPSAANSIFCSWMALGNILUGSGSNTNNHH 209
       |||| | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      128 GVGLLDPCGVCGVTPEALLSDFLRPDH--CRQAYSVAFAFMSLGGCGLYLLPAID--- 182

QY      210 KW-----FPFLKTSACEACANUKAPLVAVVFLVLCVTLTFIAKEVPYRANENLP TK 264
Db      183 -WDTSAPLYLGTQECC-----LFG--LUTLIPT-CVAATLLVADEEAAALGTE--PABG 231

QY      265 AGGEVETPTGPLAVLKGFKDL-----PGGMPSVLLVTATTLSLWSPFFILY 310
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      232 LSAPSLSPHCCPCRARLAFNIGCALLPRHLQLCRCRPRIRLRFVALCSWMAIMTFTFL 291

QY      311 DTDMMGREIYHDPK---GSNAQISAFNEGVRVGAFLGLLNLSVILGFSSFLEPMCKRVG 367
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      292 YTDVFGBGLYGQVPRAEPGTEAR-RHYDEGRVMGSLGLFLOCAISLVFSLVMDRLVOREG 350

QY      368 PRVVWTSTNFNMVCVMAATALISFWSLRDYHGYVDQAITANASIKAVCLVLFALCVPLA 427
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      351 TRAVILASVAFPVAGATCL-----SHSVAVVTASAALTGFTFSALOILPYTLA 400

QY      428 ILY 430
       ||
Db      401 SLY 403

RESULT 12
us-09-605-785-708
; Sequence 708, Application us/09605785
; Patent No. 6321716

```

CORRESPONDENCE ADDRESS:
 ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
 STREET: 90 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: U.S.A.
 ZIP: 10016
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
 MEDIUM TYPE: DISKETTE
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/595,553A
 FILING DATE: FEBRUARY 1, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: CRAIG J. ARNOLD
 REGISTRATION NUMBER: 34,287
 REFERENCE/DOCKET NUMBER: 96700/393
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 697-5995
 TELEFAX: (212) 286-0854 or 286-0082
 TELEX: TWX 710-581-4766
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 618
 TYPE: AMINO ACID
 TOPOLOGY: LINEAR
 MOLECULE TYPE:
 DESCRIPTION: PROTEIN
 HYPOTHETICAL: YES
 ORIGINAL SOURCE:
 ORGANISM: RAT
 INDIVIDUAL ISOLATE: SODIUM/IODIDE SYMPORTER
 US-08-595-553A-2

Query Match 4.6%; Score 126.5; DB 4; Length 618;
 Best Local Similarity 21.7%; Pred. No. 0.00043;
 Matches 121; Conservative 71; Mismatches 210; Indels 155; Gaps 25;

QY 1 MARG-----DGGQLAELSGVGAAGAAVVDHVPISLGRLLIAGWAGVGYG--- 47
 DB 36 LARGGORSADFFTGGRQAAVPEVGLSLAASFMSAVQ-----VLGVPAEARYGLKF 87
 QY 48 -W-ALQL-SLLT-----PYVOTLGLSHA-----LTSPMMLCGTAGLVVQPLVGLY 90
 DB 88 LWCAGQLNLSLLTAFLELPFYRGLTSTYQYLELRFSAVRLCTQYLV----- 140
 QY 91 SDRCTSRWGRRRPFIITGCMILICAVI---VVGFSDDIGAALGDTKEHCSLYHGPRWHA 147
 DB 141 -----TMLYTGIVIVAPALILNQVTGL-DIWASLSLTGIICTLTVTYVGMKA 186
 QY 148 IYV-----VLGFLLDFSNNTVQPARAMMA-----DLCDDHGPSAANSIFCS 190
 DB 187 VVMTDVFQVVMVLGVFWILARGVILLGGPRNVLSLAQNHRSINLMDDFDPPRSPRYTFT 246
 QY 191 WMAIGNILYSSGSGTNNHKKFPFLKTSACCACANLKGAFV--AVFVLVLCVTLIF 248
 DB 247 FIVGGTLVWLSMYGVNQAO-----VORYVACHTEGAKALLVNLQGLFLIVASAACCGI 301
 QY 249 AKEVPYRANENLPTTKAGGEVETEPTGPLAVLKGFKDLPPGMPSVLLVTAITWLSWFFPI 308
 DB 302 VMFVYKDCDPLLTGRISA---PDQYMLPLVLDIFEDL-PGVPGLFLACA----- 347
 QY 309 LYDTDMGRIYHGDPKGNSNAQISAFNEGVRVGAFLNLSNVILGFSSFLPEMCRKVG 368
 DB 348 -----YSGTISTASTSINAM-AAVTV-----EDLIKPRMPGLAP 380
 QY 369 R-VVWVTSNF-----MVCVMAATAALI-----SEWLSRDYHGVQDAITANASIKAV 414
 DB 381 RKLIVISKLSFIYGSACLTVAALSSLLGGVGLQSGFTVMGVISGPLLCAFTLGMLLPA- 439

QY 415 CLVLFALFLGVPLAILYSVPFAVTAQL-----AATRGGGGGLCTGVNLINISIVPQVIALGA 470
 DB 440 CNTPCVLSGLAAGLAVSLWAVGATLYPPGEQTMGLVLTSAAGCTNDSVLL----- 490
 QY 471 GPWDALFGKGNIPAFGV 487
 DB 491 GPPGATNASNGIPSSGM 507
 RESULT 14
 5268463-9
 ; Patent No. 5268463
 ; APPLICANT: JEFFERSON, RICHARD A.
 ; TITLE OF INVENTION: PLAANT PROMOTER a-GLUCURONIDASE GENE
 ; CONSTRUCT
 ; NUMBER OF SEQUENCES: 9
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/447,976
 ; FILING DATE: 08-DEC-1989
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 119,102
 ; FILING DATE: 10-NOV-1987
 ; APPLICATION NUMBER: 264,586
 ; FILING DATE: 31-OCT-1988
 ; SEQ ID NO:9
 ; LENGTH: 435
 5268463-9

Query Match 4.2%; Score 114; DB 6; Length 435;
 Best Local Similarity 19.0%; Pred. No. 0.0046;
 Matches 92; Conservative 61; Mismatches 161; Indels 170; Gaps 18;

QY 43 GVOYGMALQLSLLTPYVOTLGLSHALTSMFLCGLPIAGLVVQPLVGLYSDRCTSRWGR 102
 DB 17 GIVY-----MYLMYYTDDVGLSVGLVCTFLVARIMDAINDPVGIVNATRSRWGRKF 71
 QY 103 PFILTCM-----LICVAVIVVG-----FSSDIGAALGD 131
 DB 72 PWILIGTLANSVILLFLFSAHLFEQTOIVFVCTVYILWGMTYTIMDIPFWSLVPTILD 131
 QY 132 TREHCSLYHGPRWHAIV-YVLGFLLDFSNNTVQPARAMMADLCDHGPSAANSIFCS 190
 DB 132 KRERQLVPYPRFFASLAGFVTAGVTLPFN----- 162
 QY 191 WMAIGNILYSSGSGTNNHKKFPFLKTSACCACANLKGAFVAVVVLVLCVTLIFLFAK 250
 DB 163 -----YVGGG---DRGFGQMT-----LVLTAFIVSTIITLRNVH 196
 QY 251 EVPYRANENLPTTKAGGEVETEPTGPLAVLKGFKDL--PPGMPSVLLVTAITW----- 301
 DB 197 EV-FSSDNQPSAE-----GSHLTAKAIVAIYKNDQLSCLGLMALAYNVASNI 244
 QY 302 ---LSWFPFLLYDMDWGREIYHGDPKGNSNAQISAFNEGVRVGAFLNLSNVILGFSS 357
 DB 245 TGFALYYSYVIGDADLPYLYS-----AGANLV---TILVFFPR 283
 QY 358 LIEMCRKVGPRVVMVTSNFMVCMVMAATAALISFWSLRDHYGVQDAITANASIKAVCLV 417
 DB 284 LVKSLSR---RILWAGASILPVLSCGVLLM---ALMSYHNV--LIVAGILLNVGTA 334
 QY 418 LFAFLGVPL-----ATLYSVFFAVTAQLAATRGGGGGLCTGVNLINISIV 460
 DB 335 LFWLQVIMVADIVDGYEYKLVHRCESIAYSV-----QTMVVKGSFAFFIAVVLGMIGY 388
 QY 461 IPOV 464
 DB 389 VPV 392
 RESULT 15
 US-09-134-001C-3020
 ; Sequence 3020, Application US/09134001C
 ; Patent No. 6380370

GENERAL INFORMATION:
 APPLICANT: Lynn Doucette-Stamm et al
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 FILE REFERENCE: GTC-007
 CURRENT APPLICATION NUMBER: US/09/134.001C
 CURRENT FILING DATE: 1998-08-13
 PRIOR APPLICATION NUMBER: US 60/064,964
 PRIOR FILING DATE: 1997-11-08
 PRIOR APPLICATION NUMBER: US 60/055,779
 PRIOR FILING DATE: 1997-08-14
 NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO 3020
 LENGTH: 467
 TYPE: PRT
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3020

Query Match 4.18; Score 112; DB 4; Length 467;
 Best Local Similarity 22.08; Pred. No. 0.0081;
 Matches 115; Conservative 56; Mismatches 179; Indels 172; Gaps 25;

Qy 28 PISGLRLTLAGWAGGV-QYGMALQSLTTPYVOTLGLSHALTSFMW-LCGPIAGLVVQP 85
 Db 57 PISL-----INGFVALPMTGGIELAFSLRYVVDLSPGTPQATLLFTVYGLFAI----- 108
 Qy 86 LVGLYSDRCTRWGR---RRPFLITGCLICVAVIVVGFSSDIGAALGDTKEHCS-LYHG 141
 Db 109 -----SSWGSGLVAETFGPRIMLIGVAAWIVFVLFGLFLEQOXYTLWVYFG 158
 Qy 142 PRWHAIVVYVIGF--WLDFSNNTVQGPARAMADLCHDHGSPSAANSIFCSWMAIG-NIL 198
 Db 159 IRGLAYPLFIYAFWVWIAQVT-----PGARM-----ASAMGWFSWYICIGLL 202
 Qy 199 GYSSGSTNNHKKWPFLLKTSACCACANLKGAFLVAVVFLVLC-----TVTL 246
 Db 203 G-----NWIPSLISIRIGFINTLWGVFWVAVAGLMIMYLKRGACKPDDAVTL 252
 Qy 247 IFAKEVPYRANENLPTTKAGEVETEPTGPLAVLKGFKDLPPGMPSVLLVTAITWLSWF- 305
 Db 253 V-----ERLKLSSGVITIAER-----GMFLIVLVRILCNLSLFG 288
 Qy 306 -PFILYDPMWGREIYHGDGPKSNAQISAFNEGVRVGFAGLLNSVLGFS----- 355
 Db 289 LPVIL-----PLYLTSVEVGFSGMEQWLHWG 314
 Qy 356 -----SFLIEPMCRKVGPRVW-VTSNFMVCMVAMAATALISFWSLRDHYGVYQDAITANA 409
 Db 315 VMFVSIPTNVIWIGIGDRLCWLIQMRWFGCIGCALSSL-APYXLPQFYG-----AHF 366
 Qy 410 SIKAVCLVLFALVGLVPLAIXSVFP-AVTAQLAATRG-----GOCCLCTGVNLISIVIPQV 464
 Db 367 EI-----ALIAAIGGISVTAFPVPGAVFLALAPEQKGAASAHNLAAGLSNF----- 414
 Qy 465 IIALGAPWDALFGKGNTPAFGVSASFA--LVGGVGVVFLLP 504
 Db 415 -----MGPGIATLFIATLIGKGVVMIYAGLVAGAVLTFPFI 451

RESULT 16
 US-08-403-852D-21
 Sequence 21, Application US/08403852D
 Patent No. 5891695
 GENERAL INFORMATION:
 APPLICANT: Blanc, Veronique
 APPLICANT: Blanche, Francis
 APPLICANT: Crouzet, Joel
 APPLICANT: Jacques, Nathalie
 APPLICANT: Lacroix, Patricia
 APPLICANT: Thibaut, Denis
 APPLICANT: Zagorec, Monique
 APPLICANT: Debussche, Laurent
 APPLICANT: De Crecy-Lagard, Valerie

TITLE OF INVENTION: Polypeptides Involved In The
 TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
 NUMBER OF SEQUENCES: 43
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
 STREET: 1300 I Street, N.W., Suite 700
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3315
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC Compatible
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA: US/08/403,852D
 FILING DATE: 10-MAY-1995
 PRIOR APPLICATION DATA: PCT/FR 93/00923
 FILING DATE: 25-SEP-1993
 PRIOR APPLICATION DATA: FR 92/11441
 FILING DATE: 25-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Meyers, Kenneth J.
 REGISTRATION NUMBER: 25,146
 REFERENCE/DOCKET NUMBER: 03806.0054-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 528 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-403-852D-21

Query Match 4.18; Score 112; DB 2; Length 528;
 Best Local Similarity 20.18; Pred. No. 0.0097;
 Matches 103; Conservative 62; Mismatches 149; Indels 198; Gaps 27;

Qy 91 SDRCTSRGRRRPFLLTGCMILICVAVIV-----GFSSDIGAALGDTKEHCSL 138
 Db 28 ADPRRW-----VILG--VICLAQLVLLDNTVNLVAVIPVLTDLGASTADIO----- 74
 Qy 139 YHGPWHAAIYVYVILGFWLLDFSNNTVQGPARAMADLCHDHGSPSAANSIFCSWMAIGNL 198
 Db 75 -----WMIN-AYALVQSGLLLTAGSLADRYGRKRL-----LMLGLVL 110
 Qy 199 GYSSGSTNNHKKWPFLLKTS-----CCEACANLKGAFLVAVVFLVLCVTLIFAKEVPYR 255
 Db 111 -FGAGSA-----WAAFAQDSQAIIAARAGMGVGGALLATTTLAV-----IMQVFDDEPR 160
 Qy 256 ANENLPTTKAGEVETEPTGPL-----AVLKG-----KDLPPGMPSVLLVTAITWLSWF 305
 Db 161 AIGLWGAASSLGFA-----AGPLLGALLDHFWSIFLINLPVALLGLLAVLV----- 211
 Qy 306 PFILYDPMWGREIYHGDGPKSNAQISAFNEGVRVGFAGLLNSV-ILGFSFLIEPMCR 364
 Db 212 -----PETKNP-----EGRRPDLGLAVLSTLGMVGVVVAIIS----- 243
 Qy 365 KVGRVWVWTSNFMVCMVAMAATALISF--WSLRDYH----- 398
 Db 244 --GPEHGHTAPQVLLPAAVAAAALTAFAVRWELHTPHMLDMGFFTDRENGSPAECSF 301
 Qy 399 -----CYVODAI-----TANASKAVCL-----VLFAGLVPLAIL 429
 Db 302 GMAGSLFLLTQHLQVLGY--DALQAGLRTAPLATIIVALNLAGLAKLLAALGATARSIA 359
 Qy 430 YSVPPFAVTAQLAANTRGCG-----GLCTGVL-----NISIVIPQVIAL----- 468

Db 360 LGMTL-LAAGLSAVAVGGSPDAGYGGMLAGLLMGAGIALAMPAMATAVMSSIPPAKAG 418
Qy 469 -GAGPMDALFGKGNIPAFGVASAFALVGGVVG 499
Db 419 AGAGV-----OQTLEFGGLGVAILGAVLG 444

RESULT 17
US-08-510-646B-22
:Sequence 22, Application US/08510646B
: Patent No. 607699
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Blanc, Francis
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: APPLICANT: De Crecy-Lagard, Valerie
: TITLE OF INVENTION: Polypeptides Involved In The
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
: TITLE OF INVENTION: Coding For These Polypeptides And Their Use
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/510,646B
: FILING DATE: 03-AUG-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/403,852
: FILING DATE: 10-MAY-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR 93/00923
: FILING DATE: 25-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 92/11441
: FILING DATE: 25-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Meyers, Kenneth J.
: REGISTRATION NUMBER: 25,146
: REFERENCE/DOCKET NUMBER: 03806.0054-01000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 528 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-510-646B-22

Query Match 4.1%; Score 112; DB 3; Length 528;
Best Local Similarity 20.1%; Pred. NO. 0.0097;
Matches 103; Conservative 62; Mismatches 149; Indels 198; Gaps 27;

Qy 91 SDRCTSRWGRRRPFLTGCMCLCAVIVV-----GFSSDIGAALGDTKEHCSL 138
Db 28 ADRPRW-----VILG-----VICLAQLVLLDNTVLNVAIPVLTDLGASTADIQ----- 74

Qy 139 YHGRWHAIVVVLGFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANSIFCSWMAIGNIL 198
Db 75 -----WMIN-AYALVQSGLLLTAGSLADRYGRKRL-----LMLGLVL 110
Qy 199 GYSGSTNNWHKWFPELKTSA---CCEACANLKGAFLVAVVFLVCLTVTLIFAKEVYR 255
Db 111 -FGAGSA-----WAAFAQDSAQLIAARAGMGVGGALLATTITLAV-----IMQVFDDEPR 160
Qy 256 ANENLPTTKAGEVETPTGPL---AVLKGPF-----KDLPPGMPSVLLYTAITWLSWF 305
Db 161 AIGLWGAASSLGFA-----AGPLLGALLDHFWMGSGIFILNLPVALLGLLAVRLV----- 211
Qy 306 PFILYDTDMGRIYHGDPKGNSAOISAFNEGVRVGAFLGLINSV-ILGFSFLIEPMCR 364
Db 212 -----PETYKNP-----EGRRPDLGLGAVLSTLGMVGVYVAILIS----- 243
Qy 365 KVGPRVWVTSNFMVVCVMAATALISF--WSLRDYH-----VLFAGLVPLAIL 429
Db 244 -GPEHGWTAPQVLLPAAVAAAALTAfVRWELHTPHMLDMGFTDRRFGNGSPAECSF 301
Qy 399 -----GYVODAI-----TANASIKAVCL-----VLFAGLVPLAIL 429
Db 302 CMAGSLFLLTQHLQVLVGY--DALQAGLRTAPLALTIVALNLAGLGAALLAALGTARSIA 359
Qy 430 YSVPEAVTAQAATRGQ-----GLCTGVL-----NISVIPOVIAL----- 468
Db 360 LGMTL-LAAGLSAVAVGGSPDAGYGGMLAGLLMGAGIALAMPAMATAVMSSIPPAKAG 418
Qy 469 -GAGPMDALFGKGNIPAFGVASAFALVGGVVG 499
Db 419 AGAGV-----OQTLEFGGLGVAILGAVLG 444

RESULT 18
US-09-231-818-21
: Sequence 21, Application US/09231818
: Patent No. 6171846
: GENERAL INFORMATION:
: APPLICANT: Blanc, Veronique
: APPLICANT: Crouzet, Joel
: APPLICANT: Jacques, Nathalie
: APPLICANT: Lacroix, Patricia
: APPLICANT: Thibaut, Denis
: APPLICANT: Zagorec, Monique
: APPLICANT: Debussche, Laurent
: APPLICANT: De Crecy-Lagard, Valerie
: TITLE OF INVENTION: Polypeptides Involved In The
: TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
: TITLE OF INVENTION: Coding For These Polypeptides And Their Use
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/231,818
: FILING DATE: 25-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/403,852
: FILING DATE: 10-MAY-1995
: APPLICATION NUMBER: PCT/FR 93/00923
: FILING DATE: 25-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/403,852
: FILING DATE: 10-MAY-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Meyers, Kenneth J.
: REGISTRATION NUMBER: 25,146
: REFERENCE/DOCKET NUMBER: 03806.0054-01000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 528 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear

APPLICATION NUMBER: FR 92/11441
 FILING DATE: 25-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Meyers, Kenneth J.
 REGISTRATION NUMBER: 25,146
 REFERENCE/DOCKET NUMBER: 03806.0054-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 528 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-09-231-818-21

Query Match 4.18; Score 112; DB 4; Length 528;
 Best Local Similarity 20.18; Pred. No. 0.0097;
 Matches 103; Conservative 62; Mismatches 149; Indels 198; Gaps 27;

QY 91 SDRCTSRWRRRRPILGCMILICVAVIV-----VFSSDICAALGDTKEHCSL 138
 DB 28 ADPRRW-----VILG--VICLAQLVLLDNTVLNVAIPVLTDLGASTADIQ----- 74
 QY 139 YHPRWHAAYVVLGFWLLDFSNNTVQPARAMMADLCHDHGPSAANSIFCSWMALGNIL 198
 DB 75 -----WMIN-AYALVQSGLLTAGSLADRYGRKRL-----LMLGLVL 110
 QY 199 GYSGSTNNWHKPPFKTSA---CCACANLKGAFVAVVFLVCLTVLIFAKEVPYR 255
 DB 111 -FGAGSA-----WAAFAQDSQAIAARAGMGVGGALLATTTTAV-----IMQVDDDERPR 160
 QY 256 ANENLPTTKAGEVETETGPL---AVLKG-----KDLPPGMPVSVLLVTAITWLSWF 305
 DB 161 AIGLWGAASSLGFA---AGPLLGALLDHFWSGISFLINLPVALLGLLAVARLV----- 211
 QY 306 PFILYTDWGMREIYHGDPKGNSAQISAFNEGVRVGFAGLLNSV-ILGFSSFLIEPMCR 364
 DB 212 -----PETKNP-----EGRRPDLGLAVSLTGMVGVVYAIIS--- 243
 QY 365 KVGPRVWVTSNFMVCMAMATALISF--WSLRDYH----- 398
 DB 244 --GPEHGWTAPOVLLPAVAALTAFAVRWELHTPHPMLODMGFTRDRFNGSPAECSSE 301
 QY 399 -----GYODAI-----TANASIKAVCL-----VLFAFLGVPLAIL 429
 DB 302 GMAGSLFLLTHQLQLVLY--DALQAGLRTAPLALTIVALLMAGLAKLLAALGTAKSIA 359
 QY 430 YSVFPAVTAQLAATRGGGQ-----GLCTGVL-----NISIVIPQVITAI 468
 DB 360 LGMTL-LAAGLSAVAVGSGPDAGYGGMLAGLLMAGIALAMPAMATAVMSSIPPAKAG 418
 QY 469 -GAGPDALFGKGNIPAFGVASAFALVGVGVG 499
 DB 419 AGAGV-----QGTLETFGGGLGVAILGAVLG 444

RESULT 19
 US-08-882-704A-6
 Sequence 6, Application US/08882704A
 Patent No. 5879906
 GENERAL INFORMATION:
 APPLICANT: Jefferson, Richard A.
 APPLICANT: Wilson, Katherine J.
 APPLICANT: Leader, Michael
 TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED AND BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington

COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/882,704A
 FILING DATE: 25-JUN-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: No. 5879906tenburg Ph.D., Carol
 REGISTRATION NUMBER: 39,317
 REFERENCE/DOCKET NUMBER: 190106.404
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 622-6031
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 457 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-882-704A-6

Query Match 3.98; Score 106; DB 2; Length 457;
 Best Local Similarity 20.88; Pred. No. 0.032;
 Matches 103; Conservative 58; Mismatches 205; Indels 130; Gaps 18;

QY 36 LAGWAGGVQYGMAL---QLSLLTPYVQTLGLSHALTSMMLCGPIAGLVVQPLVGLYS 92
 DB 10 IVGSLGDVANNFAFAGALFLLSVYTDVAGVGAAGTLLLVYRVEDAFADVAGRVVD 69
 QY 93 RCTSRWRRRPFIITG-CMLICVAVIVVGFSSDICAALGDTKEHCSLYHGRWHAAYV 151
 DB 70 SVNTRWKFPPFLFGTAPLMIFSVLVFWVPTDWS-----HGSKVYVAYLT 116
 QY 152 LGFWLLDFSNNTVQPARAMH-----ADLCHDHGPSAANSIFCSWMALGNILGYS 204
 DB 117 MGLGLCYSLNIPYGLSLATANTQOQSPARLGAARGTAASLTFFVCLAFGLGSIKNS 176
 QY 205 --TNNHKKWPFPLKTSACCEACANLKGAFVAVVFLVCLTVLIFAKEVPYRANENLPT 262
 DB 177 EMVSVYHFWTIVLA-----IAGMVLFCFKSTRENVRIVAQP---SLNISLOT 223
 QY 263 TKAGEVETETGPIAVLKGFKDLPPGMPVSVLLVTAITWLSWFFFLYDTHWGMREIYH 322
 DB 224 LKRRPLFMLCIGALCVL-----ISTFAVSASS-LFYVRYVLDNLTGLTFLVLVQ 272
 QY 323 DPKGNSAQISAFNEGVRVGFAGLLNSVILGFSSFLIEPMCRKVGPRVWVTSNFMV 381
 DB 273 NLVGTVASAPLVPMVARIGKKNFTLIGALLGTCGYLL-----FFWV-SVWSLPV 321
 QY 382 AMAATALIS-----FWSLR---DYHGYVODAITANASIKAVCLVLFALFGLVPLAI 428
 DB 322 ALVALATASIGQGVMTMTVMWALEADTVEYGEVL-----TGVRIGLTVSLPSF 369
 QY 429 LYSVFFAVTAQLAATRGGGQ-----CTGVNLNISIVIPQVITAIAGAPWDALFG 478
 DB 370 -----TRKQGAIGGSIAPAFILGLSGYIANQVOTPEVIN----- 403
 QY 479 KGNIPAFGVASAFALV 494
 DB 404 -----GIRTSIALV 412

RESULT 20
 US-09-151-957-6
 Sequence 6, Application US/09151957
 Patent No. 6429292
 GENERAL INFORMATION:
 APPLICANT: Jefferson, Richard A.


```

; GENERAL INFORMATION:
; APPLICANT: Aaron Kaplan et al.
; TITLE OF INVENTION: ENHANCING INORGANIC CARBON FIXATION BY
; TITLE OF INVENTION: PHOTOSYNTHETIC ORGANISMS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
; STREET: 2001 Jefferson Davis Highway, Suite 207
; CITY: Arlington
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
; COMPUTER: Twinhead slimnote-890TX
; OPERATING SYSTEM: MS DOS version 6.2.
; SOFTWARE: Word for Windows version 3.11
; SOFTWARE: Word for Windows version 2.0 converted to
; SOFTWARE: an ASCII file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/332.041
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Friedman, Mark M.
; REGISTRATION NUMBER: 33,883
; REFERENCE/DOCKET NUMBER: 325/45
; TELEPHONE: 972-3-562553
; TELEFAX: 972-3-562554
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 474
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-332-041-5

```

```

Query Match 3.8%; Score 103.5; DB 4; Length 474;
Best Local Similarity 22.1%; Pred. No. 0.059;
Matches 60; Conservative 45; Mismatches 85; Indels 81; Gaps 17;

Qy 26 VAPISLGLILAGVYA-----GGVQY---GWA---LQLSLTPPVQTLGLSHALTSEW 73
Db 2 VSPISIRSLMFGGSPQEWGRGSLVRLVGVGQSWIQASVLWPHFEALGT--ALVAITIF 59

Qy 74 LCGPIAGLVQPLVGLYSDRCTSRGRRRPFILTGC-----MLI-----CVAVIV 118
Db 60 IAAPTSTT---MLGIFMLCCAFWA-----LITFADQPGKGLTPIHVLVFAVWICISAIA 111

Qy 119 VGFSS-DIGALGDTKEHCS-----LYHGPRW---HAAVYVGLFWLDFSS--NNTV 164
Db 112 VGFSPVMAAASGLAKLTANLCLFLAARLLQNKQWLNRLVTVLLVGLLVGSGYLRQOV 171

Qy 165 QGPARAMADLCHHGCPA-----ANSIFCSWALGNIL-----GYSSGSGTNHWHWF 212
Db 172 DG-----VEQLATWNDPTSTLAQATRVY--SELGNPNLLAAYLVPMPTGLSLSALVYVRRW 225

Qy 213 PFLKTSACCEACANLKGAFLVAVVFLVCLT 243
Db 226 P-----KLLGATWIVNLLCLEFT 244

```

```

RESULT 23
US-09-031-392-2
; Sequence 2, Application US/09031392
; Patent No. 5942398
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.

```

```

; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031.392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 563 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-031-392-2

```

```

Query Match 3.7%; Score 101; DB 2; Length 563;
Best Local Similarity 19.1%; Pred. No. 0.14;
Matches 113; Conservative 78; Mismatches 214; Indels 186; Gaps 29;

Qy 4 GDGGGLAE---LSAGVGAAGVVDHVAIPISLGRLLIAGWAGGVGYGHALQ--LSLLTPV 59
Db 28 GPGRALLECDHLRSGVGGRRRKKDWSCSLLVAS--LAGAFSGSFLYGVNLSVYNAPTPI 85

Qy 60 O-----TGLSHALTSEFMCLGPAGLVQPLVGLVSDRCTSRWGR 101
Db 86 KAFYNESHERHGRPIDPDTLTLLWSVTVSIFAIGLVGLTLVKNIG-----KVLGRK 138

Qy 102 RPFILTCMLICVAVIVVGVFSSDIGAALGDTKEHCSLYHGP-----RWHAAI-----V 149
Db 139 HTLLANNGFASIAALLMA-----CSLQAGAFEMLVGRFIMGDGVVAL 182

Qy 150 YVLGFWLDFSNNTVOGPARAMADLCHDHGSPSAANSIF--CSWMAIGNILGYSS--GSTN 206
Db 183 SVLPMLYSEISPEIRSGISLQVTA-----IFICIGVFTGQLLPELLELGRS 229

Qy 207 NWHKWFPELKTSAACCEACANLKGAFLVAVVFLVCLTIVT-----LIFAKEVPYRANENL 260
Db 230 TW-----PY-----LFGVIVVAVVQLLSLPFLPDSPRYLLLEKHNARAVKAF 273

Qy 261 PTTKAGGEVETPTGPLA---VLKGFKDLPPGMPSVL-----LVTAITWLS--- 303
Db 274 QTFGLKADVSQEEVEVLAESHVORSIR-----LVSVELELLRAPYRVQWVTVIVTMACYQ 328

Qy 304 -----WFPFLYDTHMGREIYHGDPKGSAQNAQISARNEGVRVGAFAFGLLLNSVILGRSS 356
Db 329 LCGLNAIWF-----YTNISIFGKA--GIPPAKIPYVTLSTGGI-----ETLAAVFSG 372

Qy 357 FLIEPWCRRKVGPRVYVTSNFMVCMATAALISFMSLRDHYGVQDATTANASIKAVCL 416
Db 373 LVIEHLGR--PLLI---GGFGLMGLFTGTLTITL--TLQDHAPW-----PYLSIVGILA 421

Qy 417 VLFAFLGVPLAIIYVSPFAVTAQL--AATRGGGGLCTGVNLISIVIPQVITIALGAGPWD 474

```

```

Db 422 IIAFSCGP-----GGIPFILTGEFFQSOR-----PAAFIAGTVNWL 460
QY 475 ALFGKG-----NIPAFGVASAFALVGGVGVGFWLLPKISKROFRAVS 515
Db 461 SNFAVGLLFPFIQKSLDTCFLVFATICTICITGAIYLYFVLPETKNRYAEIS 511

RESULT 24
US-09-299-549-2
; Sequence 2, Application US/09299549
; Patent No. 6136547
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; TITLE OF INVENTION: GLUTEX AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,549
; FILING DATE: 26-APR-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/031,392
; FILING DATE: 26-FEB-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 563 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-09-299-549-2

Query Match 3.7%; Score 101; DB 4; Length 563;
Best Local Similarity 19.1%; Pred. No. 0.14;
Matches 113; Conservative 78; Mismatches 214; Indels 186; Gaps 29;

QY 4 GDGGQLAE---LSAGVGRGAAVVDHVPISLGRLLILAGVAGVOYGWALQ-LSLLTPYV 59
Db 28 GGRALLECDHLRSGVPGGRKKDWSCLVAS--LAGAFSGSFYLYGNLSVVNAPTPI 85
QY 60 Q-----TLGLSHALTSFMMLCGPAGLVVQPLVGLYSDRCTSRWGRR 101
Db 86 KAFYNESWERRHGRPIDPTLLWSVTVSIFAIGLVGLTLVKMG-----KVLGRK 138
QY 102 RPFILTCMLICVAVIVGVFSSDIGAALGDTKEHSLYHP-----RWHAAI-----V 149
Db 139 HTLLANNGFSAALIMA-----CSLQAGAFEMLIVGRFIMGIDGGVAL 182
QY 150 YVLGFWLLDFSNNTVGGPARAMMADLCHDHGFSAAANSIF-CSWMAIGNILGYSS--GSTN 206
Db 183 SVLPWYLSISPKEINGSLGQVTA-----IPICIGVTGQLGLPELLGKES 229
```

```

QY 207 NWHKWFPLKTSACACANLKGAFLVAVVFLVLCLTVT-----LIFAKEVYPYRANENL 260
Db 230 TW---PY-----LFGVIVPAVQQLSLPFLPDSRYLLEKHNEARAVKAF 273
QY 261 PTTKAGGEVETPTGPLA---VLKGFKDLPGCMPSVL-----LVYTAITWLS--- 303
Db 274 QTFLGKADVSGVEEVLAESHVQRSIR-----LVSVELLRAPYVRWQVTVVITMACYQ 328
QY 304 -----WPFETLYDTDMGREIYHGDPKGSNAQISAFNEGVVRGAFGLLLNSVILGFSS 356
Db 329 LCLGNAINF-----YTNSIFGKA---GIPPAKIPYVTLSTGGI-----ETLAASFSG 372
QY 357 FLIEPMCKRVGPRVYVVTNSFMVCVMAATALISFWSLRDHYGYQDAITANASIKAVCL 416
Db 373 LVIEHLGRR--PLLI---GGFGLMGLFTGTLTITL-TLQDHAPWV-----PVLSTVIGILA 421
QY 417 VLFAPLGVPLAILYSVPEAVTAQL--AATRGCGGGLCTGVNLNISIVIPQVIALGAGPWD 474
Db 422 IIAFSCGP-----GGIPFILTGEFFQSOR-----PAAFIAGTVNWL 460
QY 475 ALFGKG-----NIPAFGVASAFALVGGVGVGFWLLPKISKROFRAVS 515
Db 461 SNFAVGLLFPFIQKSLDTCFLVFATICTICITGAIYLYFVLPETKNRYAEIS 511

RESULT 25
US-09-610-417-2
; Sequence 2, Application US/09610417
; Patent No. 6346374
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Weng, Xun
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/610,417
; FILING DATE: 05-Jul-2000
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/299,549
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/072002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 563 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
; US-09-610-417-2

Query Match 3.7%; Score 101; DB 4; Length 563;
Best Local Similarity 19.1%; Pred. No. 0.14;
```

```

Matches 113; Conservative 78; Mismatches 214; Indels 186; Gaps 29;
Qy 4 GDGGGLAE---LSAGVRCAAAVDHHVAPISLGRLLLAGWAGGVQYGHALQ-LSLLTPYV 59
Db 28 GPRALLECDHLRSGVPGRRRKQWSSCLLVAS--LAGAFSSFLYGYNLSVWNPAPYI 85
Qy 60 Q-----TLGLSHALTSMWLCQPIAGLVVQPLVGLYSDRCTSRWGR 101
Db 86 KAFYNESWERRHGRPIDPDTLTLLWSVTVSIFAIGGLVGTLLVKMIG-----KVLGRK 138
Qy 102 RPFILTGMCLICVAVIVVGFSSDGAALGDKHCSLYHGP-----RWHAAI-----V 149
Db 139 HTLLANNNGFAISAALLMA-----CSLQAGAFEMLIIVGRFIMGIDGGVAL 182
Qy 150 YVLGFWLLDFSNNTVQGPARAMMADLCDHGHGSAANSIF-CSWMALGNILGYSS--GSTN 206
Db 183 SVLPWYLSLSEISPKIRSLGQVTA-----IFICIGVETGOLLGLPELLKES 229
Qy 207 NWHKWFPEPKTSACCEACANLKGAFVAVVFLVLCVTVT-----LIFAKEVPYRANENL 260
Db 230 TW-----PY-----LFGVIVVPAVVQLLSLPPLDPSRYLLEKHNEARAVKAF 273
Qy 261 PTTKAGGEVETPTGCLA---VLKGFRLDPCMPSVL-----LVTAITWLS--- 303
Db 274 QTFLGKADVSOEVEVLAEASHVQRSIR-----LVSLELLRAPYVRWQVTVIVTMACYQ 328
Qy 304 -----WFPEILYDTMMGREIVHGDPKGSNAOISAFNEGVRVGAFGLLNSVILGFSS 356
Db 329 LCGLNAIWF-----YTNSTFGKA---GIPPAKIPVTLSTGGI-----ETLAAVPSG 372
Qy 357 FLIEPMCKRVGPRVWVTSNFMVCMVAMAATALISFWSLRDYHGYYQDAITANASIKAYCL 416
Db 373 LVIEHLGRR--PLLI---GGFGLMGLRFGTLITL-TLQDHAPWV-----PYLSIVGILA 421
Qy 417 VLFAPLGVPLAILYVFPVAVTAQL--AATRCGGGGLCTGVLNLSIVIPQVILALGAGPWD 474
Db 422 IIAISFCSGP---GGIPILTGEFFQOSOR-----PAAFIIAGTVNWL 460
Qy 475 ALFGKG-----NIPAFGVASAFALVGVGVFLLPKISKROFRAVS 515
Db 461 SNFAVGLLFPFIQKSLDTCFLVATICITGAIYLYFVLPETKNTYAEIS 511

```

Search completed: April 17, 2003, 11:33:53
Job time : 23 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:30:45 ; Search time 19 Seconds
(without alignments)
2065.453 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELSGVRGAA.....VFLPKISKRQFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 288829 seqs, 7561385 residues

Total number of hits satisfying chosen parameters: 288829

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	303.5	11.1	553	9	US-09-232-880-113
2	303.5	11.1	553	9	US-10-012-896-113
3	303.5	11.1	553	9	US-09-895-793-113
4	303.5	11.1	553	9	US-09-895-814-113
5	303.5	11.1	553	10	US-09-745-288-101
6	303.5	11.1	553	10	US-09-838-785-2
7	303.5	11.1	553	10	US-09-759-143-113
8	303.5	11.1	553	10	US-09-780-669-113
9	303.5	11.1	553	10	US-09-030-606-113
10	303.5	11.1	553	10	US-09-822-827-113
11	303.5	11.1	553	10	US-09-115-453-113
12	303.5	11.1	1079	9	US-09-895-793-947
13	303.5	11.1	1079	10	US-09-822-827-947
14	234	8.6	359	9	US-09-895-793-974
15	234	8.6	359	10	US-09-822-827-974
16	217.5	8.0	371	9	US-10-012-896-708
17	217.5	8.0	371	9	US-09-895-793-708
18	217.5	8.0	371	9	US-09-895-814-708
19	217.5	8.0	371	10	US-09-759-143-708

Sequence 708, App	371	10	US-09-780-669-708	8.0	217.5	20
Sequence 708, App	371	10	US-09-822-827-708	8.0	217.5	21
Sequence 852, App	400	9	US-10-012-896-852	7.0	191.5	22
Sequence 852, App	400	9	US-09-895-793-852	7.0	191.5	23
Sequence 852, App	400	9	US-09-895-814-852	7.0	191.5	24
Sequence 852, App	400	10	US-09-759-143-852	7.0	191.5	25
Sequence 852, App	400	10	US-09-780-669-852	7.0	191.5	26
Sequence 852, App	400	10	US-09-822-827-852	7.0	191.5	27
Sequence 627, App	469	10	US-09-925-302-627	5.7	155.5	28
Sequence 124, App	470	10	US-09-734-569-124	5.6	153	29
Sequence 6, Appli	490	10	US-09-739-457-6	5.4	148.5	30
Sequence 7, Appli	490	10	US-09-739-457-7	5.4	148.5	31
Sequence 3, Appli	490	10	US-09-739-457-3	5.4	147.5	32
Sequence 140, App	454	10	US-09-729-674-140	5.3	144.5	33
Sequence 4, Appli	458	10	US-09-739-457-4	5.3	144.5	34
Sequence 2, Appli	618	10	US-09-995-007-2	4.6	126.5	35
Sequence 464, App	460	10	US-09-741-669-464	4.5	122.5	36
Sequence 6145, Ap	460	9	US-09-738-626-6145	4.4	121.5	37
Sequence 4, Appli	413	10	US-09-957-664-4	4.4	120.5	38
Sequence 6795, Ap	448	9	US-09-738-626-6795	4.4	120	39
Sequence 7001, Ap	504	9	US-09-738-626-7001	4.3	118.5	40
Sequence 6734, Ap	497	9	US-09-738-626-6734	4.2	113.5	41
Sequence 3713, Ap	1157	9	US-09-738-626-3713	4.1	111.5	42
Sequence 5145, Ap	431	10	US-09-815-242-5145	4.0	110.5	43
Sequence 4, Appli	551	9	US-10-216-355-4	4.0	110.5	44
Sequence 6848, Ap	452	9	US-09-738-626-6848	4.0	108	45
Sequence 130, App	571	10	US-09-729-674-130	4.0	107	46
Sequence 11823, A	513	10	US-09-815-242-11823	3.9	107	47
Sequence 13718, A	842	10	US-09-815-242-13718	3.9	106.5	48
Sequence 6577, Ap	444	9	US-09-738-626-6577	3.9	106	49
Sequence 6, Appli	457	9	US-10-195-158-6	3.9	106	50
Sequence 6998, Ap	459	9	US-09-738-626-6998	3.8	105	51
Sequence 6452, Ap	481	9	US-09-738-626-6452	3.8	105	52
Sequence 5, Appli	474	9	US-09-828-173-5	3.8	103.5	53
Sequence 3, Appli	474	10	US-09-887-038-5	3.8	103.5	54
Sequence 3641, Ap	424	9	US-09-738-626-3641	3.8	102.5	55
Sequence 5, Appli	528	9	US-10-024-623-5	3.8	102.5	56
Sequence 40, Appl	363	9	US-09-802-640-40	3.7	102	57
Sequence 3703, Ap	491	9	US-09-738-626-3703	3.7	101.5	58
Sequence 2, Appli	563	9	US-09-981-947A-2	3.7	101	59
Sequence 22, Appl	451	9	US-10-084-205-22	3.7	100.5	60
Sequence 22, Appl	451	10	US-09-925-637-22	3.7	100.5	61
Sequence 13831, A	466	10	US-09-815-242-13831	3.7	100.5	62
Sequence 3588, Ap	483	9	US-09-738-626-3588	3.7	100	63
Sequence 10, Appl	486	9	US-10-051-902-10	3.7	100	64
Sequence 10, Appl	486	9	US-10-051-909-10	3.7	100	65
Sequence 10, Appl	727	9	US-09-815-923-10	3.6	99.5	66
Sequence 10, Appl	492	9	US-09-978-295A-7	3.6	99	67
Sequence 7, Appli	492	9	US-09-978-697-7	3.6	99	68
Sequence 7, Appli	492	9	US-09-978-192A-7	3.6	99	69
Sequence 7, Appli	492	9	US-09-999-832A-7	3.6	99	70
Sequence 7, Appli	492	9	US-09-978-189-7	3.6	99	71
Sequence 7, Appli	492	9	US-09-978-608A-7	3.6	99	72
Sequence 7, Appli	492	9	US-09-978-191A-7	3.6	99	73
Sequence 7, Appli	492	9	US-09-978-403A-7	3.6	99	74
Sequence 7, Appli	492	9	US-09-978-564A-7	3.6	99	75
Sequence 7, Appli	492	9	US-09-978-585A-7	3.6	99	76
Sequence 7, Appli	492	9	US-10-017-081A-7	3.6	99	77
Sequence 7, Appli	492	9	US-09-978-824-7	3.6	99	78
Sequence 7, Appli	492	9	US-09-981-915A-7	3.6	99	79
Sequence 7, Appli	492	9	US-09-999-833A-7	3.6	99	80
Sequence 7, Appli	492	9	US-10-167-749-7	3.6	99	81
Sequence 7, Appli	492	9	US-09-918-585A-7	3.6	99	82
Sequence 7, Appli	492	9	US-09-978-423A-7	3.6	99	83
Sequence 7, Appli	492	9	US-10-013-921A-7	3.6	99	84
Sequence 4431, Ap	551	9	US-09-738-626-4431	3.6	99	85
Sequence 1872, Ap	284	10	US-09-867-550-1872	3.6	98	86
Sequence 122, Ap	53	10	US-09-734-569-122	3.6	97	87
Sequence 4539, Ap	359	9	US-09-738-626-4539	3.6	97	88
Sequence 3714, Ap	661	9	US-09-738-626-3714	3.6	97	89
Sequence 5, Appli	671	10	US-09-843-598-5	3.6	97	90
Sequence 7, Appli	671	10	US-09-843-598-7	3.6	97	91
Sequence 1, Appli	2502	10	US-09-772-316-1	3.6	97	92

93 96.5 3.5 453 9 US-09-738-626-4026 Sequence 4026, Ap
94 96 3.5 539 9 US-10-051-902-26 Sequence 26, Appl
95 96 3.5 539 9 US-10-051-902-26 Sequence 26, Appl
96 95.5 3.5 472 10 US-09-822-863-4 Sequence 4, Appl
97 95.5 3.5 834 10 US-09-815-242-10076 Sequence 10076, A
98 95.5 3.5 933 10 US-09-815-242-11817 Sequence 11817, A
99 94.5 3.5 488 9 US-10-170-528-5 Sequence 5, Appl
100 94.5 3.5 488 9 US-10-062-960B-4 Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-232-880-113
Sequence 113, Application US/09232880
Publication No. US20020182596A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer Lynn
TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF
TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.428C6
CURRENT APPLICATION NUMBER: US/09/232.880
CURRENT FILING DATE: 1999-01-15
NUMBER OF SEQ ID NOS: 338
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 113
LENGTH: 553
TYPE: PRT
ORGANISM: Homo sapien
US-09-232-880-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;
QY 33 RLILAGVAGVQYQWALQSLTTPYVOTLGLSHALTSMMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLTFTGLEVCCLAAGITVPPLLEVGVEEFMTVMVLGIGVLCVPLIGSASD 76
QY 93 RCTSRWRRRRPILTCGMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPWRHAAIVYVL 152
DB 77 HWRGRYGRRRPFITWALSGLLSLFLIPRAGWLGLL-----CP---DPRPLELALLIL 127
QY 153 GFWLLDFSNTVQGPARAMADLC---DHHGSPSAANSIFCSWMALGNITLGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPTLEALLSDFLRDPDH--CRQAYSVAFWMSLGGCGLYLLPAID--- 182
QY 210 KW-----FPFLKTSACCEACANLKGAFIVAVVFLVCLTTLTIFAKEVPPYRANENLPTTK 264
DB 183 -WDTSAALPYLTQTEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231
QY 265 AGGEVETEPTGPLAVLKGFKDL-----PPCMPSVLVLTVAITWLSWPPFIFY 310
DB 232 LSAPSLSPHCCPCRARLAFRNLGALLPRLHQCCRPRTLRLFAELCSWMALMTFTLF 291
QY 311 DTDWGMREIYHGDPK---GSAQAISAFNEGVRVGFAGLLNSVILGFSFLIEPCMKRVG 367
DB 292 YTFDVEGEGYGVVPAEPGTEAR-RHYDEGVRMGSGLFLQCAISLVFLVMDRLVQRF 350
QY 368 PRVWVTSNFMVCMVMAAATALISFWSLRDYGIVODATITANASIKAVCLVLFAGLVPLA 427
DB 351 TRAVYLASVAAPVPAAGATCL-----SHSVAVVTASALTGFTFSALQILPYTLA 400
QY 428 ILY 430
DB 401 SLV 403

RESULT 2
US-10-012-896-113

Sequence 113, Application US/10012896
Publication No. US20020183251A1
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqiu
APPLICANT: Kalos, Michael D.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepner, William T.
APPLICANT: Henderson, Robert A.
APPLICANT: Hural, John
APPLICANT: McNeill, Patricia D.
APPLICANT: Houghton, Raymond L.
APPLICANT: Vinals de Bassols, Carlota
APPLICANT: Foy, Teresa
APPLICANT: Fanger, Gary R.
APPLICANT: Wantanabe, Yoshihiro
APPLICANT: Meagher, Madeleine Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C27
CURRENT APPLICATION NUMBER: US/10/012.896
CURRENT FILING DATE: 2001-12-10
NUMBER OF SEQ ID NOS: 1011
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 113
LENGTH: 553
TYPE: PRT
ORGANISM: Homo sapiens
US-10-012-896-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;
QY 33 RLILAGVAGVQYQWALQSLTTPYVOTLGLSHALTSMMLCGPIAGLVQPLVGLYSD 92
DB 17 QLLLVNLTFTGLEVCCLAAGITVPPLLEVGVEEFMTVMVLGIGVLCVPLIGSASD 76
QY 93 RCTSRWRRRRPILTCGMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPWRHAAIVYVL 152
DB 77 HWRGRYGRRRPFITWALSGLLSLFLIPRAGWLGLL-----CP---DPRPLELALLIL 127
QY 153 GFWLLDFSNTVQGPARAMADLC---DHHGSPSAANSIFCSWMALGNITLGYSSGSTNNWH 209
DB 128 GVLGLDFCGQVCFPTLEALLSDFLRDPDH--CRQAYSVAFWMSLGGCGLYLLPAID--- 182
QY 210 KW-----FPFLKTSACCEACANLKGAFIVAVVFLVCLTTLTIFAKEVPPYRANENLPTTK 264
DB 183 -WDTSAALPYLTQTEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231
QY 265 AGGEVETEPTGPLAVLKGFKDL-----PPCMPSVLVLTVAITWLSWPPFIFY 310
DB 232 LSAPSLSPHCCPCRARLAFRNLGALLPRLHQCCRPRTLRLFAELCSWMALMTFTLF 291
QY 311 DTDWGMREIYHGDPK---GSAQAISAFNEGVRVGFAGLLNSVILGFSFLIEPCMKRVG 367
DB 292 YTFDVEGEGYGVVPAEPGTEAR-RHYDEGVRMGSGLFLQCAISLVFLVMDRLVQRF 350
QY 368 PRVWVTSNFMVCMVMAAATALISFWSLRDYGIVODATITANASIKAVCLVLFAGLVPLA 427
DB 351 TRAVYLASVAAPVPAAGATCL-----SHSVAVVTASALTGFTFSALQILPYTLA 400
QY 428 ILY 430

```
Db 401 SLY 403
||
RESULT 3
US-09-895-793-113
; Sequence 113, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895.793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-895-793-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLIIAGWAGVQVGNALQLSLTPYVOTLGLSHALTFSFWMLCGPAGLVVQPLVGLYS 92
Db 17 QLLLVNLLTFGLVCLAAAGITVPPLLEVGVEEKFTMWLGIGVPLGLVCVPLLGSA 76
Qy 93 RCTSRWRRRRFFILTCGMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIVV 152
Db 77 HWRGRRRRFFILTCGMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIVV 127
Qy 153 GFWLLDFSNNTVQGPARAMADLC---DHHGPSAANSIFCSWMALGNITLGYSSGSTNN 209
Db 128 GVLLDFCGQVCFPLEALLSDLRDPDH--CRQAYSVAAFMISLGGCLGLVLLPAID 182
Qy 210 KW-----PPFLKTSACCACANLKGAFVAVVFLVCLTVTLIPAKVPPYRANENLPT 264
Db 183 -WDTSAAPYLGTQEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPT--PAEG 231
Qy 265 AGGEVEEPTGPLAVLKGKDL-----PGMPSVLLVTAITWLSNFPFLY 310
Db 232 LSAPSLPHCCPCRRARLAFRLNGLALLPRHLQCCMPRTLRLRFLVAELCSWMALMT 291
Qy 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVVRGAFGLLLNSVLGFSFSLIEPMCKRV 367
Db 292 YTFVGEGLYGVRAEPGTAR-RHYDEGVMSGLGLFLQALSLVPSLVMDVLRVRF 350
Qy 368 PRVWVTSNFMVCMATAALISFWSLRDHYGVQDAITANASIKAVCLVLAFLGVPLA 427

Db 351 TRAVYLASVAAPVAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
||
Qy 428 ILY 430
||
Db 401 SLY 403
||
RESULT 4
US-09-895-814-113
; Sequence 113, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895.814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-895-814-113

Query Match 11.1%; Score 303.5; DB 9; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLIIAGWAGVQVGNALQLSLTPYVOTLGLSHALTFSFWMLCGPAGLVVQPLVGLYS 92
Db 17 QLLLVNLLTFGLVCLAAAGITVPPLLEVGVEEKFTMWLGIGVPLGLVCVPLLGSA 76
Qy 93 RCTSRWRRRRFFILTCGMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIVV 152
Db 77 HWRGRRRRFFILTCGMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIVV 127
Qy 153 GFWLLDFSNNTVQGPARAMADLC---DHHGPSAANSIFCSWMALGNITLGYSSGSTNN 209
Db 128 GVLLDFCGQVCFPLEALLSDLRDPDH--CRQAYSVAAFMISLGGCLGLVLLPAID 182
Qy 210 KW-----PPFLKTSACCACANLKGAFVAVVFLVCLTVTLIPAKVPPYRANENLPT 264
Db 183 -WDTSAAPYLGTQEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPT--PAEG 231
Qy 265 AGGEVEEPTGPLAVLKGKDL-----PGMPSVLLVTAITWLSNFPFLY 310
Db 232 LSAPSLPHCCPCRRARLAFRLNGLALLPRHLQCCMPRTLRLRFLVAELCSWMALMT 291
Qy 311 DTDWNGRIYHGDPK---GSNAQISAFNEGVVRGAFGLLLNSVLGFSFSLIEPMCKRV 367
```

[illegible]

```

; Patent No. US20020009455A1
; GENERAL INFORMATION:
; APPLICANT: Lau, Ted
; APPLICANT: Lin, Rick
; APPLICANT: Parkes, Debbie
; APPLICANT: Parry, Gordon
; APPLICANT: Schneider, Douglas
; APPLICANT: Steinbrecher, Renate
; APPLICANT: Van Heut, Pam T
; APPLICANT: Wu, John
; TITLE OF INVENTION: DNA Encoding a No. US20020009455A1el PROST 03
; FILE REFERENCE: 51831AUSM1
; CURRENT APPLICATION NUMBER: US/09/838,785
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/200,065
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-838-785-2

Query Match      11.1%; Score 303.5; DB 10; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps

QY 33 RLILAGVAGVQYQWALQISLLTPYQTLGLSHALTSFNMWLCGPTAGLVVQPLVGLYSD 92
Db 17 QLLLVNLLTFGLVCLAAAGITYPPPLLELVGVBEKFMWVLGIPVLGLVCVPLLGASD 76
QY 93 RCTSRWGRRRPFILTCGMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVYVL 152
Db 77 HWRGRYGRRRPFILWLSGLILLSLFLIPRAGWLAGLL-----CP---DPRPLELALLIL 127
QY 153 GFWLDDFSNNTVQGPARAMMADLC---DHGHPSAANSIFCSWMAIGNILGYSSGSTNNHH 209
Db 128 GVGILLDCGGVCVCTPTEALLSDLFRDPDH--CRQAVSVYAFMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPELKTSAACEACANLKGAFLLVAVFLVCLTTLTIFAKEVPYRANENLPTTK 264
Db 183 -WDTSAAPYLGTEQC-----LFG--LLTLIFLT-CVAATLLVAEEAALGPTG--PAEG 231
QY 265 AGGEVETPTGPLAVLKGFDL-----PPGMPSVLLVTAITWLSWFFPILY 310
Db 232 LSAPLSLPHCCPCRARLAFRNIGALLPRLHQLCRMPTLRLRFLVAELCSMMALMTFTLF 291
QY 311 DTDMMGREIYHGDPK-----GSAOIASAFNEGVRVGAFGLLNSVILGFSSFLIEPMCRKVG 367
Db 292 YTDVFGEGYGGVPRABPGTEAR--RHYDEGVRMGLGLFQCAISLVFSLVMDRLVQRFG 350
QY 368 PRVWVNTSNFMVVCVMAATALISFWSLRDHYGYVQDQAITANASIKAVCLVLFAGLVPLA 427
Db 351 TRAVYLASVAAPVPAAGATCL-----SHSVAVVTAASAALTGFTFSALQILPYTLA 4700
QY 428 ILY 430
Db 401 SLY 403

RESULT 7
US-09-759-143-113
; Sequence 113, Application US/09759143
; Patent No. US2002002248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqul
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.

```

```

; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-759-143-113

Query Match      11.1%; Score 303.5; DB 10; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGVAGGVQYGMALQLSLLTPYVOTLGLSHALTSMMLCGPIAGLVQPLVGLYSD 92
Db 17 QLLVNLTLFTGLEVCCLAAGITVPPLLEVGVEEFMTVLGIGPVGLVCVPLLGASD 76
Qy 93 RCTSWGRRRPFILTCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVYVL 152
Db 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLGLL-----CP---DPRPLELALLIL 127
Qy 153 GFWLDFNSNTVQGPARAMMADLC---DHGSPSAANSIFCSMMALGNILGYSSGSTNNWH 209
Db 128 VGLLDFCGQVCFPTLEALLSDFRDPDH--CRQAYSVYAFMISLGGCLGYLLPAID--- 182
Qy 210 KW-----PFELKTSACCACANLKGAFLVAVVFLVCLTITLIFAKEVPYRANENLPTK 264
Db 183 -WDTSALAPYLGTQEEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGTE--PAEG 231
Qy 265 AGGEVETEPTGLAVLKGFKDL-----PPGMPSVLLVTAITWLSWFPFFILY 310
Db 232 LSAPLSLPHCCPCRARLAFRNLGALLPRHLQCCRMPTLRLFLVAELCSNMALMTFTLF 291
Qy 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGAFGLLLNSVILGFSFLIEPMCKRVG 367
Db 292 YTFVYEGGLYQGVPRAPETEAR-RHYDEGVRMGSLGLFLOCAISLVFSLVMDRLVQRF 350
Qy 368 PRVWVNTSNFMVVCVMAATALISFWSLDRYHGYVODAITANASIKAVCLVLPFAFLGVPLA 427
Db 351 TRAVYLVAAFPVPAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
Qy 428 ILY 430
Db 401 SLY 403

RESULT 8
US-09-780-669-113
; Sequence 113, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
```

```

; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-780-669-113

Query Match      11.1%; Score 303.5; DB 10; Length 553;
Best Local Similarity 25.3%; Pred. No. 5.4e-18;
Matches 107; Conservative 76; Mismatches 179; Indels 61; Gaps 14;

Qy 33 RLILAGVAGGVQYGMALQLSLLTPYVOTLGLSHALTSMMLCGPIAGLVQPLVGLYSD 92
Db 17 QLLVNLTLFTGLEVCCLAAGITVPPLLEVGVEEFMTVLGIGPVGLVCVPLLGASD 76
Qy 93 RCTSWGRRRPFILTCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRHAAIVYVL 152
Db 77 HWRGRRRRPFILWALSGLLSLFLIPRAGWLGLL-----CP---DPRPLELALLIL 127
Qy 153 GFWLDFNSNTVQGPARAMMADLC---DHGSPSAANSIFCSMMALGNILGYSSGSTNNWH 209
Db 128 VGLLDFCGQVCFPTLEALLSDFRDPDH--CRQAYSVYAFMISLGGCLGYLLPAID--- 182
Qy 210 KW-----PFELKTSACCACANLKGAFLVAVVFLVCLTITLIFAKEVPYRANENLPTK 264
Db 183 -WDTSALAPYLGTQEEC-----LFG--LLTLIFLT-CVAATLLVAEEAALGTE--PAEG 231
Qy 265 AGGEVETEPTGLAVLKGFKDL-----PPGMPSVLLVTAITWLSWFPFFILY 310
Db 232 LSAPLSLPHCCPCRARLAFRNLGALLPRHLQCCRMPTLRLFLVAELCSNMALMTFTLF 291
Qy 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGAFGLLLNSVILGFSFLIEPMCKRVG 367
Db 292 YTFVYEGGLYQGVPRAPETEAR-RHYDEGVRMGSLGLFLOCAISLVFSLVMDRLVQRF 350
Qy 368 PRVWVNTSNFMVVCVMAATALISFWSLDRYHGYVODAITANASIKAVCLVLPFAFLGVPLA 427
Db 351 TRAVYLVAAFPVPAAGATCL-----SHSVAVVTASAALTGFTFSALQILPYTLA 400
Qy 428 ILY 430
Db 401 SLY 403

RESULT 9
US-09-030-606-113
; Sequence 113, Application US/09030606
; Patent No. US20020081580A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF PROSTATE CANCER AND METHOD
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
```



```
Db 654 GVLGDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAEMISLGGCLGYLLPAID--- 708
QY 210 KW-----FPFLKTSACCEACANLKGAFVAVVFLVCLTTLTIFAKEVPIYRANENLPTTK 264
Db 709 -WDTSLAPYLGTQEEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 757
QY 265 AGGEVETEPTGTLAVLKGFKDL-----PGMPSVLLVLTATITWLSWPPFIY 310
Db 758 LSAPSLSPHCCPCRARLAFRNLGALLPRHLQOLCCRMPTLRLFLVAELCSNMALMTFTLF 817
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCRKVG 367
Db 818 YTFDFVGEGLYQGVPRAPGTEAR-RHYDEGVRMGLGFLQCAISLVFSLVMDRLVQRFG 876
QY 368 PRVWVWTSNFMVCVMAATALLISFWSLRDYGCVODATITANASIKAVCLVLFALFGLVPLA 427
Db 877 TRAVILASVAAPFVAAGATCL-----SHSVAVVTASAAALGTFTFSALQILPYTLA 926
QY 428 ILY 430
Db 927 SLV 929

RESULT 14
US-09-895-793-974
; Sequence 974, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 974
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-974

Query Match 8.6%; Score 234; DB 9; Length 359;
Best Local Similarity 23.9%; Pred. No. 2.9e-12;
Matches 91; Conservative 69; Mismatches 147; Indels 74; Gaps 16;

QY 33 RLILAGWAGGVQYQWALQSLTTPYVQTLGLSHALTSMWLCGPIAGLVQPLVGLXSD 92
Db 17 QLLLVNLTFFGLEVCLAAGITVPPLLLEVGVEEKFMTWLVGIGVPLGVCVPLIGSASD 76
QY 93 RCTSRWGRRRPILTGCMCLICVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVL 152
Db 77 HWRGRYGRRRPFIWALSGLLSLFLIPRAGWLAGLL-----CP--DPRPLELALLIL 127
QY 153 GFWLLDFSNNTVQGPARAMADLC---DHHGPSAANSIFCSNMALGNITLGYSSGSTNNWH 209
Db 128 GVLGDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAEMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCEACANLKGAFVAVVFLVCLTTLTIFAKEVPIYRANENLPTTK 264
Db 183 -WDTSLAPYLGTQEEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231
QY 265 AGGEVETEPTGTLAVLKGFKDL-----PGMPSVLLVLTATITWLSWPPFIY 310
Db 232 LSAPSLSPHCCPCRARLAFRNLGALLPRHLQOLCCRMPTLRLFLVAELCSNMALMTFTLF 291
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCRKVG 367
Db 292 YTFDFVGEGLYQGVPRAPGTEAR-RHYDEGVRMGLGFLQCAISLVFSLVMDRLVQRFG 876
QY 368 PRVWVWTSNFMVCVMAATALLISFWSLRDYGCVODATITANASIKAVCLVLFALFGLVPLA 427
Db 877 TRAVILASVAAPFVAAGATCL-----SHSVAVVTASAAALGTFTFSALQILPYTLA 926
QY 428 ILY 430
Db 927 SLV 929
```

```
Db 77 HWRGRYGRRRPFIWALSGLLSLFLIPRAGWLAGLL-----CP--DPRPLELALLIL 127
QY 153 GFWLLDFSNNTVQGPARAMADLC---DHHGPSAANSIFCSNMALGNITLGYSSGSTNNWH 209
Db 128 GVLGDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAEMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCEACANLKGAFVAVVFLVCLTTLTIFAKEVPIYRANENLPTTK 264
Db 183 -WDTSLAPYLGTQEEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231
QY 265 AGGEVETEPTGTLAVLKGFKDL-----PGMPSVLLVLTATITWLSWPPFIY 310
Db 232 LSAPSLSPHCCPCRARLAFRNLGALLPRHLQOLCCRMPTLRLFLVAELCSNMALMTFTLF 291
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCRKVG 367
Db 292 YTFDFVGEGLYQGVPRAPGTEAR-RHYDEGVRMGLGFLQCAISLVFSLVMDRLVQRFG 876
QY 368 PRVWVWTSNFMVCVMAATALLISFWSLRDYGCVODATITANASIKAVCLVLFALFGLVPLA 427
Db 877 TRAVILASVAAPFVAAGATCL-----SHSVAVVTASAAALGTFTFSALQILPYTLA 926
QY 428 ILY 430
Db 927 SLV 929

RESULT 15
US-09-822-827-974
; Sequence 974, Application US/09822827
; Patent No. US20020081680A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C1
; CURRENT APPLICATION NUMBER: US/09/822,827
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 974
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-822-827-974

Query Match 8.6%; Score 234; DB 10; Length 359;
Best Local Similarity 23.9%; Pred. No. 2.9e-12;
Matches 91; Conservative 69; Mismatches 147; Indels 74; Gaps 16;

QY 33 RLILAGWAGGVQYQWALQSLTTPYVQTLGLSHALTSMWLCGPIAGLVQPLVGLXSD 92
Db 17 QLLLVNLTFFGLEVCLAAGITVPPLLLEVGVEEKFMTWLVGIGVPLGVCVPLIGSASD 76
QY 93 RCTSRWGRRRPILTGCMCLICVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVL 152
Db 77 HWRGRYGRRRPFIWALSGLLSLFLIPRAGWLAGLL-----CP--DPRPLELALLIL 127
QY 153 GFWLLDFSNNTVQGPARAMADLC---DHHGPSAANSIFCSNMALGNITLGYSSGSTNNWH 209
Db 128 GVLGDFCGQVCFPLEALLSDFRDPDH--CRQAYSVAEMISLGGCLGYLLPAID--- 182
QY 210 KW-----FPFLKTSACCEACANLKGAFVAVVFLVCLTTLTIFAKEVPIYRANENLPTTK 264
Db 183 -WDTSLAPYLGTQEEC-----LFG--LLTLFLT-CVAATLLVAEEAALGPTG--PAEG 231
QY 265 AGGEVETEPTGTLAVLKGFKDL-----PGMPSVLLVLTATITWLSWPPFIY 310
Db 232 LSAPSLSPHCCPCRARLAFRNLGALLPRHLQOLCCRMPTLRLFLVAELCSNMALMTFTLF 291
QY 311 DTDWNGREIYHGDPK---GSNAQISAFNEGVRVGFAGLLNSVILGFSFLIEPMCRKVG 367
Db 292 YTFDFVGEGLYQGVPRAPGTEAR-RHYDEGVRMGLGFLQCAISLVFSLVMDRLVQRFG 876
QY 368 PRVWVWTSNFMVCVMAATALLISFWSLRDYGCVODATITANASIKAVCLVLFALFGLVPLA 427
Db 877 TRAVILASVAAPFVAAGATCL-----SHSVAVVTASAAALGTFTFSALQILPYTLA 926
QY 428 ILY 430
Db 927 SLV 929
```

RESULT 16

US-10-012-896-708
; Sequence 708, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012.896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 708
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-708

Query Match 8.0%; Score 217.5; DB 9; Length 371;
Best Local Similarity 26.7%; Pred. No. 8e-11;
Matches 99; Conservative 52; Mismatches 147; Indels 73; Gaps 19;
Qy 55 LTPYVOTLGLSHAL---TSEFMWLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFILTGCM 111
Db 10 LVPLPLALYLSQPLTHTTSLLAGIGPVLGVCVPLGSGASDHWGRGRRRPFIWALSIG 69
Qy 112 ICVAVIVVGFSSDICAALGDTKEHCSLYHGPRWHAIVVYVGLFWLLDFSNNTVQGP 171
Db 70 ILLSLFLPRAGWLAGLL-----CP---DPRPLELALLILGVGLLDFCGQVCFTPLEAL 120
Qy 172 MADLC---DHHGPSAANSIFCSWMAALGNILGYSSGSTNNHWK-----FPFLKTSACCEA 223
Db 121 LSDLRDPDH--CROAYSVYAFMISLGGCLGYLLLPDID----WDTSALAPYLGTQEEC--- 172
Qy 224 CANLKGAFLVAVVFLVLCGLTTLIFAKEVPYRANENLPTTKAGGEVETEPTGPLAVLKG 283
Db 173 ---LFG--LLTLIFLT-CVAATLLVAEEAALGPTE--PAEGLSAPLSLPHCCPCRLARLAF 224
Qy 284 KDL-----PPGMPSVLLVTAITWLSWFPFLYDTDMGRIYHGDPK---G 326
Db 225 RNIGALLPRLHQLCCMRPRTLRRLEVAELCSWMLMTFTLFYDFVGEGLYQGVPRABPG 284
Qy 327 SNAQISAFNEGVRVCA---FGLLLNSVILGFSSFLIEPM-----CRKVGPRVV---- 371
Db 285 TEAR--RHYDEGKALAASRGWCGSRPPTTLGAVSGLV-PLHPGPDFSVRKVGMDPDI 342
Qy 372 --WVTSNFMVC 380

Db 343 FSWVN-NISAC 352

RESULT 17

US-09-895-793-708
; Sequence 708, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 708
; LENGTH: 371
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-895-793-708

Query Match 8.0%; Score 217.5; DB 9; Length 371;
Best Local Similarity 26.7%; Pred. No. 8e-11;
Matches 99; Conservative 52; Mismatches 147; Indels 73; Gaps 19;
Qy 55 LTPYVOTLGLSHAL---TSEFMWLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFILTGCM 111
Db 10 LVPLPLALYLSQPLTHTTSLLAGIGPVLGVCVPLGSGASDHWGRGRRRPFIWALSIG 69
Qy 112 ICVAVIVVGFSSDICAALGDTKEHCSLYHGPRWHAIVVYVGLFWLLDFSNNTVQGP 171
Db 70 ILLSLFLPRAGWLAGLL-----CP---DPRPLELALLILGVGLLDFCGQVCFTPLEAL 120
Qy 172 MADLC---DHHGPSAANSIFCSWMAALGNILGYSSGSTNNHWK-----FPFLKTSACCEA 223
Db 121 LSDLRDPDH--CROAYSVYAFMISLGGCLGYLLLPDID----WDTSALAPYLGTQEEC--- 172
Qy 224 CANLKGAFLVAVVFLVLCGLTTLIFAKEVPYRANENLPTTKAGGEVETEPTGPLAVLKG 283
Db 173 ---LFG--LLTLIFLT-CVAATLLVAEEAALGPTE--PAEGLSAPLSLPHCCPCRLARLAF 224
Qy 284 KDL-----PPGMPSVLLVTAITWLSWFPFLYDTDMGRIYHGDPK---G 326
Db 225 RNIGALLPRLHQLCCMRPRTLRRLEVAELCSWMLMTFTLFYDFVGEGLYQGVPRABPG 284
Qy 327 SNAQISAFNEGVRVCA---FGLLLNSVILGFSSFLIEPM-----CRKVGPRVV---- 371
Db 285 TEAR--RHYDEGKALAASRGWCGSRPPTTLGAVSGLV-PLHPGPDFSVRKVGMDPDI 342
Qy 372 --WVTSNFMVC 380


```
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 852
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-852

Query Match      7.0%; Score 191.5; DB 9; Length 400;
Best Local Similarity 24.7%; Pred. No. 1.5e-08;
Matches 71; Conservative 49; Mismatches 120; Indels 47; Gaps 11;

QY 52 LSLTPYVQTGLSHALTSMMLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFILTCML 111
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 138 ITVPPLLEVGVEEKFTMTVLGICVLGVCPLIGSASDHWGRYGRRRPFILWALSIG 197
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 112 ICVAVIVVGFSSDGAALGDTKEHCSLYHGRWHAIVVVLGFWLLDFSNNTVQGPARAM 171
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 198 ILLSLFLIPRAGWLAGLL-----CP---DPRPLELALLILGVLLDFCGQVCFPLEAL 248
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 172 MADLC---DHHGPSAANSIFCSNMALGNILGYSSGSTNNHKKW-----FPFLKTSACCEA 223
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 249 LSDLFRDPDH--CROAYSVYAFMISLGGCLGYLLPAID----WDTSAAPYLGTOEEC-- 300
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 224 CANLKAGFLVAVVFLVCLTTLTIFAKEVPYRANENLPTTKAGGEVEEPTGPLAVLKGF 283
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 301 ---LFG--LLTLFLT-CVAATLVAEEAALGPTGTE--PAEGLSAPLSLPHCCPCRARLAF 352
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 284 KDL-----PPGMSVLLVTAITWLSWFPFPLYDWDWG 316
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 353 RNUGALLPRLHQLCCMPRTLRLFLVAELCSNMALMTFTLFYTFDVG 399
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 24
US-09-895-814-852
; Sequence 852, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 852
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens

; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 852
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-852

Query Match      7.0%; Score 191.5; DB 9; Length 400;
Best Local Similarity 24.7%; Pred. No. 1.5e-08;
Matches 71; Conservative 49; Mismatches 120; Indels 47; Gaps 11;

QY 52 LSLTPYVQTGLSHALTSMMLCGPIAGLVVQPLVGLYSDRCTSRWGRRRPFILTCML 111
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 138 ITVPPLLEVGVEEKFTMTVLGICVLGVCPLIGSASDHWGRYGRRRPFILWALSIG 197
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 112 ICVAVIVVGFSSDGAALGDTKEHCSLYHGRWHAIVVVLGFWLLDFSNNTVQGPARAM 171
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 198 ILLSLFLIPRAGWLAGLL-----CP---DPRPLELALLILGVLLDFCGQVCFPLEAL 248
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 172 MADLC---DHHGPSAANSIFCSNMALGNILGYSSGSTNNHKKW-----FPFLKTSACCEA 223
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 249 LSDLFRDPDH--CROAYSVYAFMISLGGCLGYLLPAID----WDTSAAPYLGTOEEC-- 300
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 224 CANLKAGFLVAVVFLVCLTTLTIFAKEVPYRANENLPTTKAGGEVEEPTGPLAVLKGF 283
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 301 ---LFG--LLTLFLT-CVAATLVAEEAALGPTGTE--PAEGLSAPLSLPHCCPCRARLAF 352
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
QY 284 KDL-----PPGMSVLLVTAITWLSWFPFPLYDWDWG 316
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB 353 RNUGALLPRLHQLCCMPRTLRLFLVAELCSNMALMTFTLFYTFDVG 399
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

RESULT 23
US-09-895-793-852
; Sequence 852, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.534C2
```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:30:00 ; Search time 22 Seconds
(without alignments)
2267.898 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELSGVGRGAA.....VFLPKISKRFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2317.5	84.9	537	2 T02982	probable sucrose t
2	1523	55.8	594	2 G84441	probable sucrose-p
3	1144	41.9	533	2 S43142	sucrose transport
4	1126	41.2	501	2 T14339	sucrose-proton tra
5	1116	40.9	510	2 S38557	sucrose transport
6	1112.5	40.7	515	2 T14340	sucrose-proton tra
7	1110	40.6	523	2 T12198	sucrose transport
8	1101.5	40.3	492	2 A84520	probable sucrose-p
9	1100.5	40.3	513	2 S38197	sucrose transport
10	1095.5	40.1	491	2 P96591	probable sucrose-p
11	1094	40.1	512	2 S38196	sucrose transport
12	1088	39.8	512	2 G86360	probable sucrose-p
13	1081	39.6	525	2 S28052	sucrose transport
14	1079	39.5	516	2 JQ2389	sucrose transport
15	1074	39.3	503	2 S52377	sucrose transport
16	1070	39.2	512	2 P96741	probable sucrose t
17	1063.5	38.9	507	2 S48789	sucrose transport
18	1043.5	38.2	523	2 S51114	sucrose-proton sym
19	996	36.5	474	2 A86234	hypothetical prote
20	883.5	32.4	428	2 S48788	sucrose transport
21	349	12.8	553	2 T38541	probable sucrose c
22	305.5	11.2	541	2 B87532	transporter, proba
23	246	9.0	451	2 E81781	probable integral
24	245	9.0	451	2 A81206	sugar transporter,
25	210.5	7.7	454	2 A75444	hypothetical prote
26	193.5	7.1	452	2 F75217	hypothetical prote
27	162.5	6.0	430	2 E75217	transporter PAB217
28	160	5.9	418	2 B87536	membrane protein,
29	148.5	5.4	490	2 JC5641	sugar transporter

orf3 bah 3'-region
drug efflux protei
tetracycline resis
probable NADH dehy
Na+/H+ antiporter
melibiose carrier
hypothetical prote
drug transport pro
probable membrane
probable efflux pr
MFS permease [drug
probable inorganic
probable nitrate/n
probable inner mem
probable membrane
probable amino aci
thyroid iodide tra
glycerol-3-phospha
probable transport
hypothetical prote
sugar,proton sympo
probable n-hydroxy
probable melibiose
probable permease
carbon starvation
probable carbon st
probable carbon st
hypothetical prote
probable transport
probable transport
probable transport
hypothetical prote
potassium uptake p
hypothetical prote
probable transcrip
probable transport
probable multidrug
hypothetical 51.0
hypothetical prote
hypothetical prote
citrate-carrier pr
Pfs enzyme II, ABC
L-arabinose transp
xyloside transport
high-affinity nitr
probable MFS trans
permeases homolog
probable MFS trans
chloramphenicol re
probable MFS trans
probable hexose tr
sodium iodide symp
hypothetical prote
sugar/Na+(H+) slmp
probable transport
major facilitator
metabolite transpo
bacteriochlorophyl
major facilitator
probable MFS trans
probable transport
probable transport
nitrogen fixation
hypothetical prote
hypothetical prote
proline permease
glucose transport
probable MFS trans

ALIGNMENTS

RESULT 1

T02982

probable sucrose transport protein - rice

C:Species: Oryza sativa (rice)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000

C:Accession: T02982

C:Hit: T. Imaizumi, N.; Scofield, G.N.; Furbank, R.T.; Ohsugi, R.

Plant Cell Physiol. 38, 1389-1396, 1997

A:Title: cDNA cloning and tissue specific expression of a gene for sucrose transporter

A:Reference number: Z14809; MUID:98182940; PMID:9522469

A:Accession: T02982

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-537 <H1>

A:Cross-references: EMBL:D87819; NID:q2723470; PIDN:BAA24071.1; PID:q2723471

A:Experimental source: cultivar Nipponbare, leaf

C:Genetics:

A:Gene: SUT1

C:Superfamily: common tobacco sucrose transport protein

Query Match 84.9%; Score 2317.5; DB 2; Length 537;

Best Local Similarity 81.2%; Pred. No. 2.8e-176;

Matches 435; Conservative 39; Mismatches 43; Indels 19; Gaps 3;

QY 1 MARGD-----GGQAEASGVGR-----AAAVDHDVAPISIGRLILAGWAG 42

DB 1 MARGSGAGGGGGGGLSVGVGGGARGGGGGAATAPIISIGRLILSGWAG 60

QY 43 GVOYGWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGLYSDRCTSRWGR 102

DB 61 GVOYGWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGLYSDRCTSRWGR 120

QY 103 PRILGCMILCVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIVYVGLWLLDFSN 162

DB 121 PYLTGCVLICVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIVYVGLWLLDFSN 180

QY 163 TVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHKKWFFPKTSACCE 222

DB 181 TVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHKKWFFPKTSACCE 240

QY 223 ACANLKGAFVAVLVCLVTLIFAKEVPYRANENLPTTKAGGEVETPTGLPLAVLKG 282

DB 241 ACANLKGAFVAVLVCLVTLIFAKEVPYRANENLPTTKAGGEVETPTGLPLAVLKG 299

QY 283 FKDLPPGMPVLLVTAITWLSWFFPILYDTPWGMREIYHGDPKGSNAQISAFNEGVRYGA 342

DB 300 FRLNPTGMPVLLVTAITWLSWFFPILYDTPWGMREIYHGDPKGSNAQISAFNEGVRYGA 359

QY 343 FGLLNSVILGFSFLLPEMCKRKGVPRVWVTSNPMVCVMAAATALISFWSLRDYHYGVQ 402

DB 360 FGLLNSVILGFSFLLPEMCKRKGVPRVWVTSNPMVCVMAAATALISFWSLRDYHYGVQ 419

QY 403 DAITANASIKAVCLVLAFLGVLPLAIVSVFPAVTAQALATRGQGLCTGVNLNISVIP 462

DB 420 KAITADKSIKAVCLVLAFLGVLPLAIVSVFPAVTAQALATRGQGLCTGVNLNISVIP 479

QY 463 QVIIALGAGPMDALFGKGNIPAFGVSAPALVGVGVVFLPLPKISKQFRAVSAGG 518

DB 480 QVIIALGAGPMDALFGKGNIPAFGVSAPALVGVGVVFLPLPKISKQFRAVSAGG 535

RESULT 2

G84441

probable sucrose-proton symporter [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: G84441

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: G84441

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-594 <STO>

A:Cross-references: GB:AE002093; NID:g3461813; PIDN:AAC32907.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g02860

A:Map position: 2

C:Superfamily: common tobacco sucrose transport protein

Query Match 55.8%; Score 1523; DB 2; Length 594;

Best Local Similarity 53.6%; Pred. No. 3.7e-113;

Matches 288; Conservative 82; Mismatches 117; Indels 50; Gaps 5;

QY 30 SLGRILLAGMVAGVQVQGWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGL 89

DB 59 SLVTLVLSCTVAAGVQVQGWALQSLTTPYVOTGLSHALTSMFWMLCGPIAGLVQPLVGL 118

QY 90 YSDRCTSRWGRRRPFIITGCMILCVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIV 149

DB 119 WSDRCTSRWGRRRPFIITGCMILCVAVIVWGFSSDIGAALGDTKEHCSLYHGPRWHAIV 178

QY 150 YVLGFWLLDFSNNTVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHKK 209

DB 179 FIIGFWLLDLANNVQGGPARAMADLCHDHGPGSAANSIFCSWMALGNILGYSSTGNHKK 238

QY 210 KWFFPLKTSACCEACANLKGAFVAVLVCLVTLIFAKEVPYRANENLPTTKAGGEVETPT 258

DB 239 EWFPELTSRACCAACGNLKAFLVAVLTLCTLTITVIFAKEIPTTSNKPTRIODSAPLL 298

QY 259 -----NLPTTAGG-----EVETEPTGLAV-----LKG 282

DB 299 DDLSKGLHSLKLNNGTANGIKYERVDTEQFGNSENSEHEDTYYVDGPGSVLNNLTS 358

QY 283 FKDLPPGMPVLLVTAITWLSWFFPILYDTPWGMREIYHGDPKGSNAQISAFNEGVRYGA 342

DB 359 LRHLPPAMHSVLIVNALTWLSWFFPILYDTPWGMREIYHGDPKGSNAQISAFNEGVRYGA 418

QY 343 FGLLNSVILGFSFLLPEMCKRKGVPRVWVTSNPMVCVMAAATALISFWSLRDYHYGVQ 402

DB 419 LGLLNSVILGFSFLLPEMCKRKGVPRVWVTSNPMVCVMAAATALISFWSLRDYHYGVQ 478

QY 403 DAITANASIKAVCLVLAFLGVLPLAIVSVFPAVTAQALATRGQGLCTGVNLNISVIP 462

DB 479 YIMRGNETTRTAIVVIFALLGFLPLAIVSVFPAVTAQALATRGQGLCTGVNLNISVIP 538

QY 463 QVIIALGAGPMDALFGKGNIPAFGVSAPALVGVGVVFLPLPKISKQFRAVSAGG 519

DB 539 QMIVSLGAGPMDQLFGGGLPAFVLAFAAGVIALQRLPTLSS-SFK--STGFH 592

RESULT 3

S43142

sucrose transport protein - castor bean

N:Alternate names: sucrose carrier

C:Species: Ricinus communis (castor bean)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-May-2000

C:Accession: S43142

R:Weig, A.; Komor, E.

submitted to the EMBL Data Library, March 1994

A:Description: A sucrose carrier from Ricinus communis.

A:Reference number: S43142

A:Accession: S43142

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-533 <WEI>

A:Cross-references: EMBL:Z31561; NID:g468561; PID:g468562

C:Superfamily: common tobacco sucrose transport protein

Query Match 41.9%; Score 1144; DB 2; Length 533;

Db		: :: :: : :: :: : :: :: : :: :: :	356
298	FDDWMGREYV-GETQHKAPELAVIYNKSVSAGALGLMLNSVLGFSGLGVQYYMARALG		
Qy	367 GPRVVWVTSNFMVCVMAAATALISFWSLRDYGHYVDAI--TANASIKAVCLVLFAFLGV	424	
Db		: :: : : : : : : : :	
357	GVKRLMGVWFILAILCICMTIVIT--KVASHHRPYSNGVLQTPESSVKTCALVWFSALCI	414	
Qy	425 PLATLYSPFAVTQAQLAATRGGCGCTGVLNTSIYIPQVITIALGAGPWDLFGKGNIPLA	484	
Db		:: :: : :: :: : :: :: : :: :: :	
415	PLAITFVPFPALASIVSTTTGSCQGLSLGLVNLIAYIPOMIVSVASGPMDAMFGGNLPA	474	
Qy	485 FGVAFAFALGVGVGVFLLPKISKRFRAVS-AGGH	519	
Db		: : : : : : : : :	
475	FVGAFAAAASGIFAFTMLFPSPPAEIKNLNVAGGH	510	
RESULT 6			
Tl4340	sucrose-proton transport protein - carrot		
N;	Alternate names: sucrose/H+ symporter protein		
C;	Species: Daucus carota (carrot)		
C;	Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000		
C;	Accession: Tl4340		

```

QY      500  VFLLPKISKRFRAVSAGG 518
      : |||| ||
Db      494  IVLLPRPSKDAASKLSLG 512

RESULT 7
T12198
sucrose transport protein - fava bean
C:Species: Vicia faba (fava bean)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Jun-2000
C:Accession: T12198
R:Weber, H.; Borisjuk, L.; Heim, U.; Sauer, N.; Wobus, U.
Plant Cell 9, 895-908, 1997
A:Title: A role for sugar transporters during seed development: molecular
A:Reference number: Z17451; MUID:97355984; PMID:9212465
A:Accession: T12198
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-523 <WEB>
A:Cross-references: EMBL:Z93774
C:Genetics:
A:Gene: sut
C:Superfamily: common tobacco sucrose transport protein

```

Query Match	40.7%;	Score 1112.5;	DB 2;	Length 515;
Best Local Similarity	44.9%;	Pred. No. 1.3e-80;		
Matches 224;	Conservative 97;	Mismatches 155;	Indels 23;	Gaps
Qy	28	PISLGRLLIAGVAGGVOYGWALQLSLLIPYVOTLGLSHALTSMWLCPGIAGLVVQPLV	87	
Db	29	PTATWKLVVAAATGAAGVQFGWALQLSLTPYVOLLGPIGHKWAAYIWCCGPISGMLVQPLV	88	
Qy	88	GLXSDRCTSRWGRRRRPIITGCMLICIVAVITVGFSSDIGAALGDTKEHCLSHYHGRPWHAA	147	
Db	89	GYTSDHCQSSFGRRRPIASGACGVAISVILGFADISYKAGD---DMSKTLKPR--AV	143	
Qy	148	IVYVLGFWLLDFSNNTVQGPARAMADLC--DHHGPSAANSIFCSMNALGNILGYSSGST	205	
Db	144	TVFVIGFWILDVANNNLQGPCRALLADLCSDGDTFRMRSANAFSFPFMAVGNILGVAAGSY	203	
Qy	206	NNHHKWFPPFKTSACCACANLKGAFVLVAVFVLVCLTVTLTFAKVPVRANENLPTTHKA	265	
Db	204	NNLYKLFPPFKTHACDLNLCANLKSCTPIISIALLIITTVVALSVVRE-----NSGPPDDA	257	
Qy	266	GGEVETEPTCLAV---LKGFKDLPPGQPSVLLVTAITWLSWFFPILLYDTCWMGREIYH	321	
Db	258	DAAEPPSSGKPIVFGELLGALDKLPRPMLLLGIIVTCLNNIAWFPIILFDTWMMGREIY	316	
Qy	322	GDPKGSNAQISAFNEGVRVGAFLNLSNVILGFSSFLIEPIMCRKV--GPRVWVWTSNFMVC	380	
Db	317	--GGTAGQCKLYDQVRAGALGLLNSVVLGLTSTAVEVLYRGVCGVKLLNGCFVNFILA	373	
Qy	381	VAMAATALISFWSLRD--YHGYVDATATANASKAVCLVLFALFGLVPLATLYSVPPFAVTAQ	439	
Db	374	IGLVMTVVWSKVAHQHRESANGQLPPSAGYVKAGALSLSILGIPLSITYSIPFALASI	433	
Qy	440	LAATRGGGQGLCTGVNLNISIVIPQVITIALGAGPMDALFGKGNIPAGVASAFALVGGVVG	499	
Db	434	YSSGAGAGQSLGVNLNLAIVPQMTIVSNVLAGPDSFLFGGMLPAPVWGAISSAATSGVILA	493	

[illegible]

RESULT 11

S38196
sucrose transport protein SUC2 - Arabidopsis thaliana
N:Alternate names: sucrose-proton symporter SUC2
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C:Accession: S38196; T00773
R:Sauer, N.K.
submitted to the EMBL Data Library, October 1993
A:Description: SUC1 and SUC2: two sucrose transporters from Arabidopsis thaliana.
A:Reference number: S38196
A:Accession: S38196
A:Molecule type: mRNA
A:Residues: 1-512 <SAU>
A:Cross-references: EMBL:X5382; NID:g407091; PID:g407092
R:Vysotskaya, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Oji, O.; Kwan, A.; Liu, S.; Li, R.;
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, J.
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome 1 BAC t22J18 sequence, complete sequence
A:Reference number: Z14202
A:Accession: T00773
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-490, 'A', 492-512 <VSY>
A:Cross-references: EMBL:AC003979; NID:g3172156; PID:g3287687; GSPDB:GN000059; ATSP:T22J18
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: SUC2; ATSP:T22J18.12
A:Map position: 1
A:Introns: 419/3; 441/1; 455/3
C:Superfamily: common tobacco sucrose transport protein
C:Keywords: sugar transport

Qy 12 LSAGVRGAAAVDHHVA-----PISLGRLLIAGMVGAGVQYGWALQLSLLTPYVOTLGLSHA 67
Db 6 MEKAAAGASALETTQTGELDQPERLKRLLTISSYSSAAAGVQFGWALQLSLLTPYVOLLGPHK 65
Qy 68 LTSPWMLCGPIAGLVYQPLVGLYSDRCTSRWGRRRRPFILLTGCMLICVAVIVVGFSSDIGA 127
Db 66 WASLIWLCGPISGMLVQPIVGYHSDRCTSRFGRRRPFIVAGAGLVTVAVFLIGYAADIGH 125
Qy 128 ALGDTKEHCISLYHGPWRHAAIVVVLGFWLLDFSNNTVQGPARAMMADLCDHHGPS--AAAN 185
Db 126 SMGD-----QLDPPKPTRATAIFALGFHWLLDVANNTLQGPCRAFLADLSAGNAKKRTAN 180
Qy 186 SIFCSWMALNIIYGSSGTSNNHKKWPFLLKTSACCEACANLKGAFIVAVVFLVLCITVT 245
Db 181 AFSFFMAVGNVLGYAAGSYRNLYKVVPFTWTESCDLYCANLKTCFLLSTLLLIIVFVS- 240
Qy 246 LIFAKEVYPYRANENLPTTKAGGEVEETPTGPLAVLGFKDLPPGMSVLLVTTAITWLSWF 305
Db 241 LCYVYKRPW-----TPEPTADGRASNVPFFG-EIFGAFKELKRPMMLLITVTLNIAWF 294
Qy 306 PFILYDFTDWMGRETYHG--DPKGSNAQISAFNEGVRVGAFLGLLNSVILGFFSFLTEPMC 363
Db 295 PFLUFDFTDWMGREVYNGNSDATATAASKKLYNDGVRAGALGLMLNAVLGFNSLGVWEIG 354
Qy 364 RKV-GRPVVWVTSMFMVCVMAATALISFWS---LRDYGHYVDADITANASIKAVCLVLF 419
Db 355 RKLGAARLNGIYVNFILAICLAIVTVVYTKQAEHNRD-HGGAKTGPPGN-VTAGALTFL 411
Qy 420 AFLGVPIALYSPFAVTAQLATRRGGQGLCTGVNLISIVIQVIALGAGPMDALFGK 479
Db 412 AILGIPQAITSFIPFALASIFSTNSGAGOGSLGLVLAIVPQMWISVGGGFDELFGG 471
Qy 480 GNIPAFGVASAFALVGWGVFLLP 504
Db 472 GNIPAFVILGAIAAAVSGVLTLVP 496

RESULT 12
G86360
probable sucrose-proton symporter SUC2 protein T22J18.12 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2000
C:Accession: G86360
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; Whit-
chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khayki-
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun-
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Map position: 1
C:Superfamily: common tobacco sucrose transport protein
C:Genetics:
A:Residues: 1-512 <STO>
A:Cross-references: GB:AE005172; MID:g3287687; PIDN:AAC25515.1; GSPDB:GN00141

Qy	12	LSAGVGGAAVVDHVA----	PISGRLLILAGMVAGVQYQGWALQLSLLTPYVOTGLSHA	67
		:	: :	
	6	MEKAANGASALETQTGELDQPERLKIISVSSAAGVQFGWALQLSLLTPYVOLLGIPKH	65	
Db				
Qy	68	LTSPFMWLCGPIAGLVVOPVLVGLXSDRCTSRNGRRRRPFILTCGMLICVAVIIVGCFSSDIGA	127	

Db 66 WASLIWLCGPTSGMLVQPIVGYHSDRCTSRGRRRPFIVAGAGLVTVAVFLIGYAADTGH 125
Qy 128 ALGDTKEHCSLYHGRHAAITVYVGLFWLLDFSNVTQGPARAMMADLCDDHGPS--AAN 185
Db 126 SMGB-----QDKPKPTRAIAIFALGFWILVDVANTLQGPCRAFLADLSAGNAKKTATN 180
Qy 186 SIFCSWMAIGNILYSGSGTNNHKKWPFLLKTSACCACANLKGAFVAVVFLVLCITVT 245
Db 181 AFFSFFNAVGNLVGAAGSYRNLYKVPFTMTESCDLYCANLKTCCFLSITLLIVTFVS 240
Qy 246 LIFAKEVPIRANENLPTTKAGGEVETPTGPLAVLKGFKOLPPONPSVLLVTAITWLSWP 305
Db 241 LCYVYKERPW-----TPEPTADGKASNPFFG-EIFGAFKELKRPMMMLLIIVTALNWTAF 294
Qy 306 PFILYDMDMGREIYHG--DPKGSNAQISAFNEGVRVGAFLGLLNSVLITGSSFLIEPMC 363
Db 295 PFLLFDMDMGREYVGGNSDATAAASKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIG 354
Qy 364 RKV-GPRVWVWTSNFMVCMVMAATALISFWS---LRDHYGVQDAITANASIKAVCLVLF 419
Db 355 RKLGAARLWGVNFILAIACLMTVVTQKAENHRD-HGAKTGPPGN--VTAGALTLE 411
Qy 420 AFLGVPLAILYSVPFANTQAATRGQGLCTGVLTNISIVIPQVITIALGAGPMDALFGK 479
Db 412 AILGIPQAITSIPFALASIFSTNSGAGQGLSLGLVNLAIIVVQMFVSVGGGPPDELFG 471
Qy 480 GNIPAFGVASAFALVGVGVVGLLP 504
Db 472 GNIPAFVIGATAAAVSVGLATVLP 496

RESULT 13

S28052

sucrose transport protein - spinach

N;Alternate names: sucrose carrier protein; sucrose permease

C;Species: Spinacia oleracea (spinach)

C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000

C;Accession: S28052

R;Riesmeier, J.W.; Willmitzer, L.; Frommer, W.B.

EMBO J. 11, 4705-4713, 1992

A;Title: Isolation and characterization of a sucrose carrier cDNA from spinach by functi

A;Reference number: S28052; MUID:93099843; PMID:1464305

A;Accession: S28052

A;Molecule type: mRNA

A;Residues: 1-525 <RTE>

A;Cross-references: EMBL:X67125; NID:g21318; PIDN:CAA47604.1; PID:g21319

A;Superfamily: common tobacco sucrose transport protein

C;Keywords: transmembrane protein

Query Match 39.6%; Score 1081; DB 2; Length 525;
Best Local Similarity 44.9%; Pred. No. 4e-78;
Matches 227; Conservative 89; Mismatches 166; Indels 24; Gaps 10;

Qy 27 APISLGRILAGVAGVQVQGWALQLSLLTPYVOTLGLSHALTFSFWMLCGPVQPL 86
Db 31 AEATLKKLGLVASAAGVQVQGWALQLSLLTPYVOTLGLSHALTFSFWMLCGPVQPL 90
Qy 87 VGLYSDRCTSRGRRRPFILTGCMILCVAVIVGVFSSDIGNAALGDTKEHCSLYHGRHWA 146
Db 91 VGYISDRCTSRGRRRPFIAAGAAVAVAVGLIGFAADGAAGDGTGNVA---KPR--A 145
Qy 147 AIVVVLGFWLLDFSNVTQGPARAMMADLC--DHGFSAAINSIFCSNWLGNILYSGS 204
Db 146 IAVFVGVFWILDVANNLTQGPCRALLADMAAGSQTTRYANAFSFFWALGNIGCYAAGS 205
Qy 205 TNNHKKWPFLLKTSACCACANLKGAFVAVVFLVLCITVLIPTAKE-----VPIRANEN 259
Db 206 YSRLYTVTFPTKTAACDYCANLKSCEFFISITLLIVLTILALSVMVKERQITIDEIQEED 265
Qy 260 LPT--TKAGGEVETPTGPLAVLKGFKDLPPGMPSPVLLVTAITWLSWPFLLYDMDMGRE 318
Db 266 LKNRNSSGCARLPFFGQL--IGALKOLPKPMLILLVLTALNWTAFWFFLLFDMDMGKE 323

Qy 319 IYHGDPKGSNAQISAFNEGVRVGAFLGLLNSVLITGSSFLIEPMC RKV-GPRVWVWTSNF 377
Db 324 VY-----GGTVGEGKLYDQGVHAGALGLMINSVLGVKSLSTIEGLARWVGAKRLGIVNI 379
Qy 378 MVCVMAATAALI--SFWSLRDHYGYVQDAI--TANASTKAVCLVLFAPLGVPLAILYSVP 433
Db 380 ILAVCLAMTIVLTKSAEHRDHSIHMSAVPPPPAGVKGALAIFAVLGIPLAITFISIP 439
Qy 434 FAVTAQAAATRGQGLCTGVLTNISIVIPQVITIALGAGPMDALFGKGNIPAFGVASAFAL 493
Db 440 FALASIFSASGSGQGLSLGLVNLAIIVVQMFVSVTSGPMDAMFGGGLNLPFAVVGAVAA 499
Qy 494 VGVGVGVFLLPKISKRQFRVAVSAGGH 519
Db 500 ASAVLSFTLLPSPPEAKIGSGMGH 525

RESULT 14

JQ2389

sucrose transport protein - potato

C;Species: Solanum tuberosum (potato)

C;Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000

C;Accession: JQ2389; S40310

R;Riesmeier, J.W.; Hirner, B.; Frommer, W.B.

Plant Cell 5, 1591-1598, 1993

A;Title: Potato sucrose transporter expression in minor veins indicates a role in phl

A;Reference number: JQ2389; MUID:94146554; PMID:8312741

A;Accession: JQ2389

A;Molecule type: mRNA

A;Residues: 1-516 <RIE>

A;Cross-references: EMBL:X69165; NID:g439293; PIDN:CAA48915.1; PID:g439294

A;Experimental source: cv. Desiree

C;Comment: The gene encoding for this protein is highly expressed in mature leaves.

C;Superfamily: common tobacco sucrose transport protein

C;Keywords: glycoprotein; transmembrane protein

F;31-53/Domain: transmembrane #status predicted <TM1>

F;67-86/Domain: transmembrane #status predicted <TM2>

F;103-122/Domain: transmembrane #status predicted <TM3>

F;141-160/Domain: transmembrane #status predicted <TM4>

F;180-200/Domain: transmembrane #status predicted <TM5>

F;226-248/Domain: transmembrane #status predicted <TM6>

F;285-304/Domain: transmembrane #status predicted <TM7>

F;331-349/Domain: transmembrane #status predicted <TM8>

F;366-385/Domain: transmembrane #status predicted <TM9>

F;409-427/Domain: transmembrane #status predicted <TM10>

F;429-448/Domain: transmembrane #status predicted <TM11>

F;3;92/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 39.5%; Score 1079; DB 2; Length 516;
Best Local Similarity 43.9%; Pred. No. 5.7e-78;
Matches 225; Conservative 101; Mismatches 162; Indels 24; Gaps 11;

Qy 16 VRGAAAVDVHVPATISLGRLLIAGVAGVQVQGWALQLSLLTPYVOTLGLSHALTFSFWMLC 75
Db 15 VSSSLVEQVPLAPSKLIIIVASIAAGVQVQGWALQLSLLTPYVOTLGLSHALTFSFWMLC 74
Qy 76 GPIAGLVVQPLVGLYSDRCTSRGRRRPFILTGCMILCVAVIVGVFSSDIGNAALGDTKEH 135
Db 75 GPISGMIVQPVGVVGYISDNCSSRRGRRRPFIAAGAAVAVVAVLIGFADLGHASGDT--- 131
Qy 136 CSLYHGRHWAIIIVVVLGFWLLDFSNVTQGPARAMMADLCDDHGPS----AANSIFCSW 191
Db 132 --LGKGFKPRAIVAVVGVFWILDVANNLTQGPCRALLADLSG--GKSGRMTANAFFSFF 187
Qy 192 MALGNILYSGSGTNNHKKWPFLLKTSACCACANLKGAFVAVVFLVLCITVLIPTAKE-----VTILIF 248
Db 188 MAVGNILGYAAGSYSHLFKVEPFPSKTRACDMYCANLKSCEFFIA-IFLLSLTTLTALTIVR 246
Qy 249 AKEVPIRANENLPTTKAG-GEVETPTGPLAVLKGFKDLPPGMPSPVLLVTAITWLSWPF 307
Db 247 ENELPEKDEQIDBKLAGAGSKGVFFFG--EIFGALKELPRPMWILLVLTALNWTAFWFF 304

```

Qy 308 ILYDTDMGRIEYHGDPKSGNAQISAFNEGVRVGFAGFLLLSNVILGFSFLEIEMCRKV- 366
      ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 FLYDTDMAKEVFGG--QVGDARL--YDLGVVRAGAMGLLLQSWLGFMSLGVFEFLGKRG 360

Qy 367 GPRVWVTSNFMVVCVMAATALISFWSLRD-YHGYYQDAITANASIKAVCLVLFAFLGVP 425
      | : : | : : : : | : : : : | : : : : | : : : : | : : | : : | : : |
Db 361 GAKRLWGLNFVLAICLAWTILVTYKMAKSRQHDPAQTJMGTPGVKIGALLLFAALGIP 420

Qy 426 LAILYSVPFAVTAQLAATRGGGGGLCTGYLNTSIVIPQVILALGAGPDALFGKGNIPAF 485
      || : |||| : : : |||| : : : |||| : : : |||| : : : |||| : : : |||| :
Db 421 LAATESIPFALASIFSNRSGSGGLSLGVNLAIYVPMQLVSLVCGPDDLLFGGGLPGF 480

Qy 486 GVASAFALGVGVVGFLLPKIKRQPRAYSAG 517
      | : | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 481 VVGAVAAAAAVLALTMLPSPADAKPAVAG 512

RESULT 15
S52377
sucrose transport protein SUC1 - common plantain
N:Alternate names: sucrose proton symporter SUC1
C:Species: Plantago major (Common plantain)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-May-2000
C:Accession: S52377
R:Sauer, N.K.H.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52377
A:Accession: S52377
A:Molecule type: mRNA
A:Residues: 1-503 <SAU>
A:Cross-references: EMBL:X84379; NID:gi1732005; PIDN:CAA59113.1; PID:g667047
C:Genetics:
A:Gene: SUC1
C:Superfamily: common tobacco sucrose transport protein
C:Keywords: sugar transport

Query Match 39.3%; Score 1074; DB 2; Length 503;
Best Local Similarity 43.5%; Pred. No. 1.4e-77;
Matches 232; Conservative 92; Mismatches 157; Indels 52; Gaps

Qy 9 LAELSAGVRAAAVVDHVPISLGRLLIAGMAGGVQYQWALQLLSLLTPYQTLGLSHAL 68
      : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MGELSGIENGKATGEATIPV--WKEVIVASLAAGTQFGWALQLSLLTPYQTMGLPGHGA 58

Qy 69 TSFNLWCGIAGLVQVPLVGLYSDRCTSRWGRRRRPFILTCMLICVAVIVVGFSSDIAA 128
      || : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db 59 ASFTLWCGVSGLLVQPLAGYFSDRCKSRFRRRRPFINSAGCLVAANVILGFAADIGHS 118

Qy 129 LGD--TKEHCSLYHGPRWHAAYIVYLGFWLLDFSNNTVQGPARAMADLC--DHHGPSAA 184
      || : || : || : || : || : || : || : || : || : || : || : || :
Db 119 AGDDMTKK-----TKPR--AVVVFVGVFILDVANNMLQGPCRAFLADLSAGDEKKMTHA 171

Qy 185 NSTFCSHMALGNILGYSSGSTNNHKKWFFPFLKTSACCACANLKGAFVAVVFLVCLTV 244
      | : | : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db 172 MSFFAFEMGIGNVLYGAAGSYNNLRLHLPFTDACEIFCANLKCFLIHIC-LLMCLTI 230

Qy 245 TLIFAKVPPYRANENLPTTKAGGE--VETEPGTPLAVLKGFKDLPPGMPSLVLLTAVTWL 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 TALSVMKEP--LVNVYDDEKKGSLMVFVLEFGAL-----KNLSKPWMLMLVTCLNWI 282

Qy 303 SWFFFLYDPTDMGREIYHGDPKSGNAQISAFNEGVRVGFAGFLLLSNVILGFSFLEIEM 362
      || : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db 283 AWFPFLLYDPTDMGREIY-----GGKVNQSVYDMGVRAIGLMLNSVVLGTTISLLYFF 337

Qy 363 CRKV-GRPVVWVTSNFMVVCVMAATALISFWSLRDHY-----GYVODAITANASIKAV 414
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 338 SKGAKAAKTMWLGNTVILVAVGLAGTYWVS-----YHAKSVRQLGASGEALPPSEVKAS 391

Qy 415 CLVLFAFLGVPLAILYSVPFAVTAQLAA----TRGGGGLCTGYLNTSIVIPQVILALG 470
      | : || : |||| : |||| : |||| : |||| : |||| : |||| : |||| : |||| :
Db 392 ALAIFAILGIPLAVTTSVPFALAAIYCSRNTWTGAGQGLSLGVWNLSIVIPQIIVLSLS 451

Qy 471 GPWDALFGKGNIPAFGVASAFALVGGVGVFLL--PKISK--RQPRAYVSAGGH 519

```

Db 452 GPLDKAFGGNLPATIMGAVGFSVGLAFILLPAPKVDNGVNPFF-AVGCGGH 503

RESULT 16

F96741

probable sucrose transport protein F17M19.4 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: F96741

R:Theologus, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, N.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Creasy, T.H.; Dewar, N.F.; Hughes, B.; Huizlar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luos, J.S.; Maiti, R.; Marzia, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: F96741

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-512 <STO>

A:CROSS-references: GB:AE0051173; NID:g6978914; PIDN:AAF34305.1; GSPDB:GN00141

C:Genetics:

A:Gene: F17M19.4

A:Map position: 1

C:Superfamily: common tobacco sucrose transport protein

Query Match 39.2%; Score 1070; DB 2; Length 512;

Best Local Similarity 42.5%; Pred. No. 2.9e-77;

Matches 216; Conservative 96; Mismatches 166; Indels 30; Gaps 9;

Qy . 11 ELSAGVGAADVHVAPISLG-----RLILAGMAGGVQYQWALQLSLTTPYVOTLGL 64

Db 5 EAERAANNATALETQSSPEDLGOPSPURKIITSVASIAAGVQFGWALQLSLTTPYIOLGI 64

Qy . 65 SHALTSEFWLGGTAGLVVOPVLGLYSDRCTSRWGRRRPFILTCMLICVAVIVVGFSSD 124

Db 65 PHKSSVWMLCGPLSGMIVQPIVGHSDRESRGRRRPFIAAGVALVAVSVFLGFAAD 124

Qy 125 IGAALGDTKEHCISLYHGRWHAATVYVLGFWLLDFSNTVQGGPARAMMADLC--DHHGPS 182

Db 125 MGHSGD-----KLENKVRTRAIIFLTGFWFLDVANNTLQGPCRAFLADLAAGDAKKTR 179

Qy . 183 AANSIFCSWALGNILYSSGSTNNHKKWPFPLKTSACCACANLKAGFLVAVVFLVCL 242

Db 180 VANACFFFMAGVNLGYAAGSVTNLHKMFPFTMTKACDIYCANLTKCFPLSLTLLIVT 239

Qy 243 TVTLIFAKPEYRANENLPTTKAGGEVETPTGPL----AVLKGFKDLPMPSVLLVTA 298

Db 240 FSSLWYVKDKWSPQ-----GDKE-BKTSLSFFGFIQAVRHKRPMWMLLIVT 290

Qy 299 ITLWSWFPFIIYLDMDMGREIYHGDPRGSAQISAFNEGVRVGAFGILLNSVILGFSSFL 358

Db 291 INWIAWFPFIIYLDMDMGREYVCGNSDGRSKKLYDQVQAGALGLMFSIILGFVSLG 350

Qy 359 IEPCKRV-GPRVWVTSTNFMVCVMAATALISFWSLRDHYGVQDAITANAS-IKAVCL 416

Db 351 VESIGRMGGAKRLLWGVNFTLAIGLMTVLVT--KSAEHREIAGPLAGPSSGIGAGVF 408

Qy 417 VLEAFLGVPLAIIYVSPFAVTAQIAATRRGGOGICLTGVNLISIVIPQVITALCAGPDA 476

Db 409 SLFTVLGPIPLAIIYSIFALASTFSTNSGAGGSLGVLNIAICIPQMIVSFSSGPDQA 468

Qy 477 FGKNIPAFGVASAFALVGGVGVFLLP 504

Db 469 FGGNLPSEFVGAATAAVSGVIALTVLP 496

RESULT 17

S48789
sucrose transport protein - common tobacco
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 26-May-2000
C:Accession: S48789
R:Buerkle, X.Y.Z.; Frommer, W.B.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48787
A:Accession: S48789
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-507 <BUE>
A:Cross-references: EMBL:X82276; NID:g575350; PID:g575351
C:Superfamily: common tobacco sucrose transport protein

Query Match 38.9%; Score 1063.5; DB 2; Length 507;
Best Local Similarity 43.9%; Pred. No. 9.6e-157;
Matches 225; Conservative 100; Mismatches 17; Indels 31; Gaps 13;
Qy 19 AAADVHVAPIS-LGRILAGWAGVQYQWALQSLTTPYVOTLGLSHALTSPFWLGGP 77
Db 11 SSLAVEQPLPKSKLWIIWASIAAGVQWALQSLTTPYVOLLGIPKFASTWLCGP 70
Qy 78 IAGLVQPLVGLYSDRCSRGRRRPFLTGCMLICVAVVVGSSDGAALGDTKEHCS 137
Db 71 ISGMIVQPVVGYSDNCSRRFRGFTAAAGALVTIAVFLIGFAADLGHATGD-----P 125
Qy 138 LYHGRPHAAIVYVGLFWLLDFSNVTQGPARAMMADLCHDHGSPA-----ANSIFCSWMA 193
Db 126 LKSKSPRAIAVVGFWILDVANNMLOGPCRALLDL---SGKARMTSNAPFFPMA 182
Qy 194 LGNLTGSSGNTNHHWFFPKTSACCEACANLKGAFVAVFLVCLTV---TLIFAK 250
Db 183 VGNLVGAAGSYRLKTFPSKTPACDIYCANLKSCEFFIA-VLLSLTLTALTAVVREN 241
Qy 251 EYPRANENLTPTKAGGEVETPGPLAVLKGFDLPKGMPSVLLVTAITWLSWPPFLY 310
Db 242 ELPEK-DEHETDERKAGARKSVPPFG-EIFGALKDLPRPMILLVTLNINIAFPPEFLY 299
Qy 311 DTDWNGRIYHGDPKGSNAQISAFNEGVRVGAFLGILLNSVILGFSSFLIEPMCRKV-GPR 369
Db 300 DTDWMAKEVYGG--KVGDRLL--YDLGVHAGALGILLNSVILGFSSFLIEPMCRKV 355
Qy 370 VVWTSNFMVCAATAATLISFWSLR-----DYHGVQDAITAMASIKAVCLVLFPAFLGVP 425
Db 356 RLWGLNPLVAVCMALTVLVTKMAEKSRQYDAHGTL---MPTSGVKIGALTLEAVLGIP 412
Qy 426 LAILYVPFAVTAQLAATRGCGGGLCTGVNLISIVIPQVITLALGAGPMDALFGKGNIPAF 485
Db 413 LAVTSPFPALASIFSSNAGSGGGLSLGLVNLAIIVPQMLVSIAGGPWDDLFGGGLNPGF 472
Qy 486 GVASAFALGVGVGFLLPKISKRQFRAVSAGG 518
Db 473 IGVAAAAAGSILALTMLPSPADAKPATWGG 505

RESULT 18
S51114
sucrose-proton symporter - beet
C:Species: Beta vulgaris (beet)
C:Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 26-May-2000
C:Accession: S51114
R:Westram, A.; Eckhardt, U.; Frommer, W.B.; Riesmeyer, J.W.
submitted to the EMBL Data Library, January 1995
A:Description: Sequence of a sugar beet sucrose transporter cDNA.
A:Reference number: S51114
A:Accession: S51114
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-523 <WES>
A:Cross-references: EMBL:X83850; NID:g633171; PID:g633172
C:Superfamily: common tobacco sucrose transport protein

Query Match 38.2%; Score 1043.5; DB 2; Length 523;
Best Local Similarity 43.6%; Pred. No. 3.8e-75;
Matches 223; Conservative 84; Mismatches 160; Indels 45; Gaps 12;
Qy 30 SLGRILLAGWAGVQYQWALQSLTTPYVOTLGLSHALTSPFWLGGPILVQPLVGL 89
Db 35 SLKLLALVASIAAGVQWALQSLTTPYVOLLGIPHTWAPYIWLGGPISGMIVQPTVGY 94
Qy 90 YSDRCTSRWGRRRPFLTGCMLICVAVVVGSSDGAALGDTKEHCSLYHGRPHAAIV 149
Db 95 YSDRCTSKFGRRRPFLVAVGATLVGFVSLIGFAADLGHATGDPNGV-----PKPRAIV 149
Qy 150 YVLGFWLLDFSNVTQGPARAMMADLCHDHGSPA-----ANSIFCSWALNLTGSSGST 205
Db 150 FVVGFWILDVANNTLOGPCRALLDMA--AGSQAKTRYANAFSFFEMALGNIGGY-AGSY 206
Qy 206 NNWIKWFFPKTSACCEACANLKGAFVAVVFLVCLTVTLIFAKEVY-----RANENL 260
Db 207 GRLYTVFPFTHTKACDTYCANLKSCEFFISITLITLITLALSVVRERPTLDEIOEENL 266
Qy 261 PTTRAGGEVETPGPLAVLKG-----KDLPGMPSVLLVTAITWLSWPPFLYDPTD 313
Db 267 KNN-----TGGCARLFFGOLFALDKLPKMLILLVTLNINIAFPFLFDFTD 316
Qy 314 WMGREIYHGDPKGSNAQISAFNEGVRVGAFLGILLNSVILGFSSFLIEPMCRKV-GPRVW 372
Db 317 WMGREVY---GGTVGEGKADMGVHAGALGLMINSVILGIMSLGIEKLARLVGGVKRLW 372
Qy 373 VTSNFMVCAATAATLISFWSLRDYG--YVQDAI---TANASIKAVCLVLFPAFLGVP 427
Db 373 GIVNLILAVCLAMTILVT-KSAEHYRATHVPGAIGPLPPPGVKGGAALFAVLGIPLA 431
Qy 428 ILYSPFAVTAQLAATRGCGGGLCTGVNLISIVIPQVITLALGAGPMDALFGKGNIPAFGV 487
Db 432 ITFSIPFALASIFSSNAGSGGGLSLGLVNLAIIVPQMFVSVTSFGPMDALFGGGLNPAFV 491
Qy 488 ASAFALGVGVGFLLPKISKRQFRAVSAGG 519
Db 492 GAVATAASAILSFLLPPPPPEAKIGGSMGGH 523

RESULT 19
A86234
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: A86234
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: A86234
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:Cross-references: GB:AF005172; NID:g2160188; PIDN:AAB60751.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
C:Superfamily: common tobacco sucrose transport protein

Query Match 36.5%; Score 996; DB 2; Length 474;
Best Local Similarity 43.2%; Pred. No. 2e-71;
Matches 205; Conservative 79; Mismatches 135; Indels 56; Gaps 9;
Qy 34 LILAGWAGVQYQWALQSLTTPYVOTLGLSHALTSPFWLGGPILVQPLVGLYSDR 93
Db 11 LILAGWAGVQYQWALQSLTTPYVOTLGLSHALTSPFWLGGPILVQPLVGLYSDR 93

Db 501 QVL 503

RESULT 22

B87532

transporter, probable CC2283 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C;Accession: B87532

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: B87532

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-541 <STO>

A;Cross-references: GB:AE005673; NID:g13423798; PIDN:AAK24254.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2283

Query Match 11.28; Score 305.5; DB 2; Length 541;

Best Local Similarity 23.98; Pred. No. 1.6e-16;

Matches 135; Conservative 76; Mismatches 211; Indels 143; Gaps 21;

Qy 14 AGVGAADVHDVAPISLGLRLLAGWAG--GVOYGWALQSLTLPYVOTGLSHALTSF 71

Db 27 AGSTGGRRGGSGMARQLSLFQIWNMCFGFGIQLGFLQANTSRIFQSLGVQVNNHAI 86

Qy 72 MWLCGPIAGLVQPLVGLYSDRCTSRWRRPFILTCMLICVAVIVVGFSSDIGAALGD 131

Db 87 LWIAAPATGLLVQPLIGHFSDKTWGRGRRRPFYFWGAILTTALLVMPNSPTL----- 140

Qy 132 TKHCSLYHGRWHAAYIVVGLFWLLDFSNNTVGPAPAMMAD--LCOHHGPS--AANSIF 188

Db 141 -----WVAA-----AALIMDASINITPEPFAFVGDNLDPQERATGYAMOSFF 184

Qy 189 CSMWALGNILGYSSGSTNNW--HKWFFPLKTSACCEACANGLAPLAVVFLVCLTVTL 246

Db 185 IG-----LGAVFASALPMLTWNFDVANTAPAGVDPDSVRIAFYGGAGLLLAFLVTV 237

Qy 247 IFAKEVPRANENLPPTKAGEV-----ETPT-----GPLAVL----- 280

Db 238 FTTRE--YSPQLTAFKAEREIAGLGLHERPEPSVNAYIALGVGVLGALALIVMGA 295

Qy 281 -----KGKFD-----LPGMPSVLLVTAIT 300

Db 296 GLEKELYVLAGLPAFGLAGVAGARFKRIGRTDNGSEVLADVFRMPKTMQLAVQPPS 355

Qy 301 WLSWFPFTLYDTDMGREIYHGDPKGNAQISAFNEGVR--VGAFAGLLLNSVILGFSSFLI 359

Db 356 WFLGFAMWIIYTPAVA--TVHEG---AVDASKAYNEGADWGVLFVAVNGV--AALAALV 409

Qy 360 EPMCRKVGPRVYVTSNPMFVCMVAMAATALISFWSLRYHGYVQDAITANASIKAVCLVLF 419

Db 410 IPLMVKVTSSRV-----SHAVCLGLGALGLLSFLVIRD-----PGLLWI 448

Qy 420 AFLGVPLA--ILYSVPFAVTAQLAATRGGGGGLCTGVNLISIVIPQVILALGAGPWFALF 477

Db 449 GMVGVGFAWSSILSTPYSILAGALPAR--KMGVYMGIFNVFVVPQLLAA-----TLL 499

Qy 478 GKGNIAPAGVASAFALGVGVGVFL 502

Db 500 GLMLKTFEGNQSIFALVGLGALSFAL 524

RESULT 23

B81781

Probable integral membrane transport protein NMA2100 [imported] - Neisseria meningitidis

C;Species: Neisseria meningitidis

C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C;Accession: B81781

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc

; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre

Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491

A;Reference number: A81775; MUID:20222556; PMID:10761919

A;Accession: B81781

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-451 <NAP>

A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAR85316.1; PID:g738

A;Experimental source: serogroup A, strain 22491

C;Genetics:

A;Gene: NMA2100

Query Match 9.08; Score 246; DB 2; Length 451;

Best Local Similarity 22.68; Pred. No. 6.8e-12;

Matches 114; Conservative 71; Mismatches 186; Indels 134; Gaps 20;

Qy 43 GVOYGWALQSLTLPYVOTGLSHALTSFMWLCGPIAGLVQPLVGLYSDRCTSRW---- 98

Db 30 GVQTAFTLSSQMSRIFQTLGADPHSLGWFILPLAGMLVQPIGVHYSDR---TWKPRL 86

Qy 99 -GRRRPFILTCMLICVAVIVVGFSSDIGAALGDTKEHCSLYHGRWHAAYIV--VLGEWL 156

Db 87 GGRRLPYLLYGLTIAIVIMILMPSNGSGFSGF-----YASL-----AALSFALMIAL 133

Qy 157 LDFSNNTVGPAPAMMADLCOHHGPSAANSIFCSMWALGNILGYSSGSTNNWHKWFPELK 216

Db 134 LDVSSNMAQPFKMWGDMVNEEQ-----GYAIGTQS-----FLA 169

Qy 217 TSACEAC-----ANL--KG-----AFLVAVVFLVCLTVTLIFAKE----- 251

Db 170 NTGAVVAAILPFVFAV IGLANTAEGVVVPQVWVAFYVGAALLVITSAFTIFKVKNPE 229

Qy 252 --VPIRANENLPPTKAGEVETEPTGPLAVLKGFKDLPPGMPSVLLVTAITWLSWFPF-- 307

Db 230 TYARYHGIDVAANOEFKANWIELLKTAPKA-----FWTVTLVQFFCFQAFQY 275

Qy 308 -ILDYDTDMGREIYHGDPKGNAQISAFNEGVRVVGAFGLL--LNSVILGFSSFLIEPMCR 364

Db 276 MWTYSAGAIENVWHITTDASSVGYQEAQN-----WYGVLAQVSAVAIVCSFVLAKVPN 329

Qy 365 KVGPRVVMVTSNPMFVCMVAMAATALISFWSLRYHGYVQDAITANASIKAVCLVLFALGV 424

Db 330 K-----YHKAGYECCLAGALGALGFPSVFFIGNQVALV-----LSVTLLGI 368

Qy 425 PLALYSVPFAVTAQLAATRGGGGGLCTGVNLISIVIPQVILALGAGPWFALFGKGNIPA 484

Db 369 AWAGIITYPLTIVTN--ALSCKHMGTYLGLFNGSICMPQIVASLLSFVLFPMLG----- 420

Qy 485 FGVASAFALGVGV-----GVFLL 503

Db 421 -GLQATMFLVGVGVLLLGAFSVFLI 444

RESULT 24

A81206

sugar transporter, probable NMB0388 [imported] - Neisseria meningitidis (strain MC58

C;Species: Neisseria meningitidis

C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C;Accession: A81206

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.

ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A;Reference number: A81000; MUID:20175755; PMID:10710307

A;Accession: A81206

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-451 <TET>

malacines 121; conservative 63; mismatches 166; models 113; gaps 23;

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2003, 11:29:55 ; Search time 26 Seconds
(without alignments)
827.931 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELSGVRGAA.....VFLPKISKRFRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1081	39.6	525	1	STP_SPIOL
2	424	15.5	530	1	MATP_MOUSE
3	404	14.8	530	1	MATP_HUMAN
4	135	4.9	639	1	LACY_LEULA
5	134	4.9	544	1	YD74_SYNY3
6	129	4.7	479	1	YVLP_LACPE
7	127	4.7	393	1	SOTA_ERWCH
8	126.5	4.6	618	1	SL55_RAT
9	122.5	4.5	454	1	PUCR_RHOSU
10	122.5	4.5	460	1	YAGG_ECOLI
11	122.5	4.5	721	1	YJIV_ECOLI
12	122	4.5	459	1	YDJK_ECOLI
13	120.5	4.4	460	1	YICJ_ECOLI
14	120	4.4	434	1	CITL_SALTY
15	119.5	4.4	464	1	ARAE_BACSU
16	119	4.4	412	1	YICM_ECOLI
17	118.5	4.3	473	1	MOT4_CHICK
18	118.5	4.3	643	1	SL55_HUMAN
19	118	4.3	404	1	Y4XM_RHISN
20	117.5	4.3	403	1	MHPT_ECOLI
21	116.5	4.3	477	1	YPUW_RHOCA
22	116.5	4.3	522	1	STPI_ARATH
23	115	4.2	463	1	YNAJ_BACSU
24	114.5	4.2	592	1	TAT2_YEAST
25	114.5	4.2	410	1	YJIO_ECOLI
26	113.5	4.2	427	1	YDFJ_ECOLI
27	111	4.1	461	1	CSBC_BACSU
28	109.5	4.0	477	1	NOMI_PSEAE
29	109	4.0	429	1	YQ84_MYCTU
30	109	4.0	430	1	DGOT_ECOLI
31	108.5	4.0	473	1	PHDK_NOSK
32	108	4.0	469	1	MELB_ECOLI
33	107	3.9	523	1	STC_RICCO

34	106.5	3.9	393	1	TCR7_VIBAN
35	106.5	3.9	468	1	ERIC_VIBCH
36	106.5	3.9	575	1	ITR1_SCHPO
37	106.5	3.9	832	1	ATCU_SALTY
38	105	3.8	410	1	YQJV_BACSU
39	105	3.8	457	1	UIDB_ECOLI
40	105	3.8	471	1	MELB_ENTAE
41	105	3.8	587	1	LACP_KLUJA
42	104.5	3.8	391	1	YIDY_ECOLI
43	104.5	3.8	413	1	MUCK_ACICA
44	104	3.8	405	1	PMRA_LACIA
45	104	3.8	476	1	MELB_SALTY
46	104	3.8	591	1	DSDI_PSEAE
47	103.5	3.8	394	1	ARAJ_ECOLI
48	103.5	3.8	633	1	NUOL_MYCTU
49	102.5	3.8	360	1	MRAY_NEIMA
50	102.5	3.8	360	1	MRAY_NEIMB
51	102.5	3.8	394	1	EMRD_ECOLI
52	102.5	3.8	477	1	YDJI_ECOLI
53	102.5	3.8	492	1	YD19_METJA
54	102.5	3.8	615	1	UAPA_EMENI
55	102	3.7	402	1	OPDE_PSEAE
56	102	3.7	431	1	CITL_ECOLI
57	102	3.7	470	1	CYCA_ECOLI
58	102	3.7	522	1	MELB_KLEPN
59	102	3.7	522	1	STA_RICCO
60	101.5	3.7	422	1	EXUT_BACSU
61	101.5	3.7	462	1	NARU_SALTY
62	101	3.7	540	1	GTR9_HUMAN
63	101	3.7	630	1	SGA4_BOVIN
64	101	3.7	693	1	SGA3_BOVIN
65	100.5	3.7	407	1	Y661_METH
66	100.5	3.7	412	1	YWFA_BACSU
67	100.5	3.7	416	1	CHRA_PSEAE
68	100.5	3.7	546	1	Y61B_MYCPN
69	100.5	3.7	599	1	YV06_CAEEL
70	100.5	3.7	687	1	AFUB_ACTPL
71	100	3.7	389	1	YD22_DEIRA
72	100	3.7	514	1	YHK8_YEAST
73	100	3.7	620	1	SGA3_HUMAN
74	99.5	3.6	461	1	PUCR_RHOCA
75	99.5	3.6	491	1	AMPG_ECOLI
76	99	3.6	399	1	PMRA_STRPN
77	99	3.6	463	1	GUTA_BACSU
78	99	3.6	487	1	TTDT_ECOLI
79	99	3.6	492	1	ANKH_HUMAN
80	99	3.6	514	1	QACA_STRAM
81	98.5	3.6	277	1	CYST_ECOLI
82	98.5	3.6	544	1	YC9D_SCHPO
83	98.5	3.6	627	1	YHE0_YEAST
84	98.5	3.6	637	1	NU5M_STRPU
85	98.5	3.6	832	1	ATCU_SALTY
86	98	3.6	425	1	DALT_KLEPN
87	97.5	3.6	343	1	Y841_METH
88	97.5	3.6	429	1	G6PU_HUMAN
89	97.5	3.6	457	1	AROP_ECOLI
90	97.5	3.6	529	1	Y0U1_CAEEL
91	97	3.6	436	1	MNTH_DEIRA
92	97	3.6	619	1	VALL_YEAST
93	97	3.6	701	1	CSTA_ECOLI
94	96.5	3.5	332	1	YPHD_ECOLI
95	96.5	3.5	334	1	PIT_RHIME
96	96.5	3.5	367	1	BRB2_PIG
97	96.5	3.5	400	1	BWR2_BACSU
98	96.5	3.5	762	1	AVP3_HORVU
99	96	3.5	363	1	AG22_HUMAN
100	96	3.5	396	1	SOTB_ECO57

ALIGNMENTS

RESULT 1

```

STP_SPIOL
ID STP_SPIOL STANDARD; PRT; 525 AA.
AC Q03411.
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sucrose transport protein (sucrose permease) (sucrose-proton
DE symporter).
OS Spinacia oleracea (Spinach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Spinacia.
OX NCBI_TaxID=3562;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leaf;
RX MEDLINE=93099843; PubMed=1464305;
RA Riesmeyer J.W., Willmitzer L., Frommer W.B.;
RT "Isolation and characterization of a sucrose carrier cDNA from
RT spinach by functional expression in yeast.";
RL EMBO J. 11:4705-4713(1992).
CC -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF SUCROSE INTO THE
CC CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).
CC CAN ALSO TRANSPORT MALTOSE AT A LESSER RATE.
CC -!- PATHWAY: Sucrose metabolism.
CC -!- SUBCELLULAR LOCATION: Inner membrane.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X67125; CAA47604.1; -
DR PIR; S28052; S28052.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr.1.
DR TIGRfams; TIGR01301; GPH_sucrose.1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW Transmembrane; Transport; Sugar transport; Symport.
FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 38 58 1 (POTENTIAL).
FT TRANSMEM 72 92 2 (POTENTIAL).
FT TRANSMEM 107 127 3 (POTENTIAL).
FT TRANSMEM 145 165 4 (POTENTIAL).
FT TRANSMEM 184 204 5 (POTENTIAL).
FT TRANSMEM 230 250 6 (POTENTIAL).
FT TRANSMEM 295 315 7 (POTENTIAL).
FT TRANSMEM 338 358 8 (POTENTIAL).
FT TRANSMEM 373 393 9 (POTENTIAL).
FT TRANSMEM 422 442 10 (POTENTIAL).
FT TRANSMEM 455 475 11 (POTENTIAL).
FT TRANSMEM 488 508 12 (POTENTIAL).
FT DOMAIN 509 525 CYTOPLASMIC (POTENTIAL).
FT SEQUENCE 525 AA; 54992 MW; 01837A4D2C0C6C6 CRC64;
Query Match 39.6%; Score 1081; DB 1; Length 525;
Best Local Similarity 44.9%; Pred. No. 1.9e-70;
Matches 227; Conservative 89; Mismatches 166; Indels 24; Gaps 10;
QY 27 APISGLRLIAGMAGVQYQWALQSLTTPYVOTPLGLSHALTSPMWLCGGPTAGLVQPL 86
DB 31 AEATLKKGLVASVAGVQWALQSLTTPYVQVLGIPHTWAAVYIWLCPISGMIVQPL 90
QY 87 VGLYSDRCTSRGRRRPFTTCMCLICVAVIVGFSSDITGAALGDTKEHCSLYHGPRWHA 146
DB 91 VGYSDRCTSRGRRRPFTIAGAAALVAVAGVGLIGFAADIGAASGDTGNVA---KPR--A 145
QY 147 AIVYVGLFWLLDFSNVTVOGPARAMMADLC--DHGSPSAANSIFCSWMALGNILGYSSGS 204

```

```

Db 146 IAVFVVGFWILDVANNTLQGPCRALLADMAAGSQTTRYANAFFSFFMALGNIGGYAAGS 205
QY 205 TNNHKKWFPFLKTSACCACACANLKGAFVAVVFLVCLITVTLIFAKE-----VYRANEN 259
Db 206 YSRLTYVFPFTKTRACDVCYCANLKSFFISITLLIVLILALSVVVKERQITIDEIQEED 265
QY 260 LPT-TKAGGEVETEPTGPLAVLKGFKDLPMPGMPSVLLVITATLWSWFFFIYLDYDMWGRE 318
Db 266 LKNNSSGCCARLFFFGQL--IGALKDLPKPMLILLVLTALNWLAWFFFLFDYDMWGRE 323
QY 319 IYHGDPKGSNAQISAFNEGVVRGAFGLLLNSVLGFSFLIEPCMKV-GRPVVWVTSNF 377
Db 324 VY---GGTVGEGKLYDQGVHAGLGLMINSVLGVMSLSIEGLARMVGGAKRLWGIYNI 379
QY 378 MVCVAMAATALI--SFWSLRDYGCVQDAI--TANASTKAVCLVLFAPGLVPLALYVSP 433
Db 380 ILAVCLAMTVLVTSAEHRDHHIMSGAVPPPPAGVKGALAIFAVLGIPLAITFSIP 439
QY 434 FAVTAQLAATRGGGGGLCTGVINISIVIPQVITIALGAGPMDALFKGNIPAFGVSAPAF 493
Db 440 FALASIFSASSSGGLSLGLVNLAIYVPMFVSVTSGPDMFGGGLNLPFVVGVAAT 499
QY 494 VGVVGVPELLPKISKRPRAVSAGGH 519
Db 500 ASAVLSFTLLPSPPPPEAKIGSGMGH 525
RESULT 2
MATP_MOUSE
ID MATP_MOUSE STANDARD; PRT; 530 AA.
AC P58355;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma
DE antigen AIM1) (Underwhite protein).
GN MATP OR AIM1 OR UW.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ddv; TISSUE=Eye, Kidney, and Uterus;
RX MEDLINE=11372467; PubMed=11479596;
RA Fukamachi S., Shimada A., Shima A.;
RT "Mutations in the gene encoding B, a novel transporter protein, reduce
RT melanin content in medaka.";
RL Nat. Genet. 28:381-385(2001).
RN [2]
RP SEQUENCE FROM N.A. AND VARIANTS UW-DBR ASN-153 AND PRO-435.
RX MEDLINE=21473748; PubMed=11574907;
RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T.,
RA King R.A., Brilliant M.H.;
RT "Mutations in the human orthologue of the mouse underwhite gene (uw)
RT underlie a new form of oculocutaneous albinism, OCA4.";
RL Am. J. Hum. Genet. 69:981-988(2001).
CC -!- FUNCTION: Melanocyte differentiation antigen. May transport
CC substances required for melanin biosynthesis (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Melanocytes, eyes, kidney and uterus.
CC -!- DISEASE: Defects in MATP are the cause of the UW-dbr phenotype
CC that results in loss of nearly all pigmentation in the homozygous
CC state.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entitles requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC EMBL: AF360357; AK81713.1; Matp.
CC MGD; MGI:2153040; Matp.
CC Disease mutation; Albinism.
DR KW Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;
KW Disease mutation; Albinism.
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 46 66 1 (POTENTIAL).
FT DOMAIN 67 68 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 69 89 2 (POTENTIAL).
FT DOMAIN 90 105 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 106 126 3 (POTENTIAL).
FT DOMAIN 127 138 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 139 159 4 (POTENTIAL).
FT DOMAIN 160 184 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 185 205 5 (POTENTIAL).
FT DOMAIN 206 216 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 217 237 6 (POTENTIAL).
FT DOMAIN 238 318 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 319 339 7 (POTENTIAL).
FT DOMAIN 340 366 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 367 387 8 (POTENTIAL).
FT DOMAIN 388 398 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 399 419 9 (POTENTIAL).
FT DOMAIN 420 425 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 426 446 10 (POTENTIAL).
FT DOMAIN 447 477 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 478 498 11 (POTENTIAL).
FT DOMAIN 499 504 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 505 525 12 (POTENTIAL).
FT DOMAIN 526 530 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 153 153 D -> N (IN UW-DBR).
FT VARIANT 435 435 S -> P (IN UW-DBR).
SQ SEQUENCE 530 AA; 57961 MW; F4EAD07916D9FC CRC64;
Query Match 15.5%; Score 424; DB 1; Length 530;
Best Local Similarity 25.3%; Pred. No. 2.6e-23;
Matches 135; Conservative 89; Mismatches 196; Indels 114; Gaps 17;
QY 30 SLGRLIAGVAGVQVGMALQJLSLLTPYVQTGLSHALTSFWMLCPIAGLVVQPLVL 89
DB 31 STGRLLVHSMAGREFCYAEAYVTPVLLSVGLPKSLYSWMLLSPIGLLPQVYGS 90
QY 90 YSDRCTSRGRRRPFILTCGMLICVAVIVVGFSSDGAAL---GDTREHCSLYHGR--- 143
DB 91 ASDHCRARWGRRRPYILT-----LAIMML-----LGMALYNGDAVV-SALVANPQKL 138
QY 144 -WHAAYVYVGLFWLLDFSNNTVQPARAMADLCHDRHGPSAANSIFCSWMLGNILGYSS 202
DB 139 IWAISITMV-GVVLDFSDAFIDGPIKAYLFDVCSHQDKKGLHYHALFTGFGCALCYIL 197
QY 203 GSTNNHKKWPFPLKTSACCACANLKA-----FLVAVFLVLCVLTVLIFAKEVPRAN 257
DB 198 GAIDMWHL-----DLGRLLGTEFQVVFVFFSALVILLCFTHLCSIPAPLRDA 245
QY 258 ENLPTRK-----AGGEVETE-----PTGP-----LA 278
DB 246 ATDPPSQDQGSLSASGMEYGSIEKVNKGADTEQVQEWKKNKPSQSQRTSMKMS 305
QY 279 VLKGFKLPPGMPVLLVTAITLWSPPFFILYDWDWGREYHGDPRK--SNAQISAFNE 336
DB 306 LLRALVNMPSHYRLCVSHLIGWTAFLSNMLFFTDENGQIVYHGDYPGAHNSTEFLLYER 365
QY 337 GVRVGAAGLLNSVILGESSFLIEPCKRGKVPVWVTSNFMVCMVAAATLISFWSLRD 396
DB 366 GVEYCGKGLCINSVFSSVSYFQKAMYSYICLKGLY-----FM-----GYLLFGLGTG 413
QY 397 YHGVQDAITANASIKAVCLVLPFLAFGLVPLAALYSVPFAVTAQL-----A 441
DB 414 FIGLPPNYSY-----LVLCSMFGVMSSTLYVPFNFLTAIEYHREEKEKGOEAPGGP 465

QY 442 ATRGGGGLCTGVNLISIVIPVILALGAGPMDALFGKNIPAFGVASAFALVG 495
DB 466 DNQGRGKGVDCALTCWVLAQILVGGGLFLVNMAGSVVVVVI-TASAVSLIG 518
RESULT 3
ID MATP_HUMAN STANDARD; PRT; 530 AA.
AC OSUMX9; Q9BTM3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma antigen AIM1).
GN MATP OR AIM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM AIM-1A).
RC TISSUE=Melanoma;
RX MEDLINE=21115844; PubMed=1121837;
RA Harada M., Li Y.F., El-Gamil M., Rosenberg S.A., Robbins P.F.; "Use of an in vitro immunoselected tumor line to identify shared melanoma antigens recognized by HLA-A*0201-restricted T cells."; Cancer Res. 61:1089-1094(2001).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM AIM-1B).
RP TISSUE=Skin;
RC Strausberg R.; Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RL [3]
RN ALTERNATIVE SPLICING.
RP Ferro S.; Unpublished observations (NOV-2001).
RL [4]
RN DISEASE, AND VARIANT LEU-374.
RX MEDLINE=21473748; PubMed=11574907;
RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davissson M.T., King R.A., Brillant M.H.; "Mutations in the human orthologue of the mouse underwhite gene (uw) underlie a new form of oculocutaneous albinism, OCA4."; Am. J. Hum. Genet. 69:981-988(2001).
CC -1- FUNCTION: Melanocyte differentiation antigen. May transport substances required for melanin biosynthesis (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By similarity).
CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms; AIM-1a (shown here), AIM-1b and AIM-1c; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in most melanoma cell lines and melanocytes.
CC -1- DISEASE: Defects in MATP are the cause of oculocutaneous albinism type 4 (OCA4). OCA4 is characterized by hypopigmentation of skin, hair and eyes. It leads to reduced visual acuity.
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to a frameshift in position 188.
CC -1- CAUTION: The described alternatively spliced isoforms are inferred using information from ests.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC
DB EMBL: AF172849; AAD51812.1;
DB EMBL: BC003597; AA003597.1; ALT_FRAME.
DB MIM: 606202; -;
KW Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;

KW Polymorphism: Albinism; Alternative splicing.
FT DOMAIN 1 46
FT TRANSSEM 47 67
FT DOMAIN 68 68
FT TRANSSEM 69 89
FT DOMAIN 90 110
FT TRANSSEM 111 131
FT DOMAIN 132 138
FT TRANSSEM 139 159
FT DOMAIN 160 184
FT TRANSSEM 185 205
FT DOMAIN 206 216
FT TRANSSEM 217 237
FT DOMAIN 238 318
FT TRANSSEM 319 339
FT DOMAIN 340 366
FT TRANSSEM 367 387
FT DOMAIN 388 398
FT TRANSSEM 399 419
FT DOMAIN 420 425
FT TRANSSEM 426 446
FT DOMAIN 447 477
FT TRANSSEM 478 498
FT DOMAIN 499 504
FT TRANSSEM 505 525
FT DOMAIN 526 530
FT CARBOHYD 356 356
FT VARSPLIC 129 187
FT VARSPLIC 188 295
FT VARSPLIC 386 406
FT VARSPLIC 407 530
FT VARIANT 374 374
SQ SEQUENCE 530 AA; 58301 MW; F14A4BACA8AF31B CRC64;
/FTID-VAR_012162.
F -> L.
Query Match 14.8%; Score 404; DB 1; Length 530;
Best Local Similarity 24.2%; Pred. No. 7.1e-22;
Matches 128; Conservative 88; Mismatches 202; Indels 112; Gaps 16;
QY 33 RILAGMAGGVOYQWALQSLTTPVOTLGLSHALTSPMVLGPGIAGLVQPLVGLYSD 92
DB 34 RLIMISMAMGFEFCYAVEAAVTPLVLSVGLPSSLSYLVNPLSLPILGFLQPVVGSASD 93
QY 93 RCTSRWRRRRPRLT-GCMILCAVIVVGVGSSDIGAALGDTKEHGSYHGPR----WHA 147
DB 94 HCRSRWRRRRPRLTILGVNMLVGMALYNGATVVAALIAN-----PRKLVW-AI 142
QY 148 IVYVLGFWLLDFSNTVVGPARAMADLCHDHGSPSAANSIFCSWALNGLIYSSGSTNN 207
DB 143 SVTMIGVVLDFDAFDIDPIKAYLDFVCSHODKEKGLHYHALFTGFGGALYLLGAIDW 202
QY 208 WHKWPFLKTSACCEACANLKA-----FLVAVFLVLCFTVLLFAKEV----- 252
DB 203 AHL-----ELGRLLETQVFMFFTSALVLTCTVHLCSISEAPLTVAKGIP 250
QY 253 PYRANENLPTTKAG-----GEVETEPTG---PLAVLKGEK----- 284
DB 251 PQOTQDPLSDGMYEYGSIEKVNGVYNPELANOGAKNHAQOTRRAMTLKSLRAL 310
QY 285 -DLPGMPSVLVITATWLSWPPFLYDMDWREIYHGDPKG--SNAQISAFNEGVRVG 341
DB 311 VNMPPHYRLCLSHLGTAFSLNFTDFMGQIVYRGDPSYSAHNSTEFILYERGVVG 370
QY 342 AGLLLNSVILGFSFLPEPCRKVGPRVWVTSNFMVCMVAMATALI-SFWSLRDYHGY 400
DB 371 CWGFCINSVSSLYSFQKVLVSYIGLKLGYETGLGFGTGLGFPNYS----- 423
QY 401 VQDAITANASIKAVCLVFAFGLGVPLAILYSVPFAVTAQL-----AATRGG----- 446
DB 424 -----TLVCLSLFGVMSSILYTPFNLIYTHREEKEKQAPGPDNSV 469
QY 447 -CGGLCTGVNLNISIVPOVIALGAGPMDALFGKGNIPAFGVASAFALVG 495

Db 470 RGKGMDCATLCMVQLAQILVGGGLFVLTAGTGVVVVVI-TASAVALIG 518
RESULT 4
LACY_LEULA
ID LACY_LEULA STANDARD; PRT; 639 AA.
AC Q48624;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lactose permease (Lactose-proton symport) (Lactose transport protein).
GN LACS.
OS Leuconostoc lactis.
OC Plasmid pNZ63.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1246;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NZ6009;
RX MEDLINE=96209221; PubMed=8633855;
RA Vaughan E.E., David S., de Vos W.M.;
RT "The lactose transporter in Leuconostoc lactis is a new member of the
RL Lacs subfamily of galactoside-pentose-hexuronide translocators.";
CC Appl. Environ. Microbiol. 62:1574-1582(1996).
CC -!- FUNCTION: RESPONSIBLE FOR TRANSPORT OF BETA-GALACTOSIDES INTO THE
CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM),
AND ALSO FOR TRANSPORT OF HOMOLOGOUS AND HETEROLOGOUS EXCHANGE OF
BETA-GALACTOSIDES.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- DOMAIN: THE ENZYME IIA-LIKE REGION MAY SERVE A REGULATORY
FUNCTION.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
SODIUM:GALACTOSIDE SYMPORTER FAMILY (SGF).
CC -!- SIMILARITY: CONTAINS 1 PTS EIIA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed, usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC HSP; U47655; AAC44113.1; -.
DR HSP; P45618; 2GPR.
DR InterPro: IPR001927; Na/Gal_symp.
DR InterPro: IPR001127; PTS_EIIA.
DR Pfam: PF00338; PTS_EIIA_1; 1.
DR ProDom: PD002243; PTS_EIIA; 1.
DR TIGRFAMs: TIGR00792; gph; 1.
DR TIGRFAMs: TIGR00830; PTBA; 1.
DR PROSITE: PS00371; PTS_EIIA_1; 1.
DR PROSITE: PS00872; NA_GALACTOSIDE_SYMP; 1.
KW Transmembrane; Sugar transport; Symport; Phosphorylation;
Plasmid.
FT DOMAIN 1 473 SGF DOMAIN.
FT DOMAIN 474 639 EIIA DOMAIN.
FT MOD_RES 557 557 PHOSPHORYLATION (BY HPR) (BY SIMILARITY).
FT TRANSSEM 20 40 POTENTIAL.
FT TRANSSEM 59 79 POTENTIAL.
FT TRANSSEM 99 119 POTENTIAL.
FT TRANSSEM 124 144 POTENTIAL.
FT TRANSSEM 176 196 POTENTIAL.
FT TRANSSEM 207 227 POTENTIAL.
FT TRANSSEM 264 284 POTENTIAL.
FT TRANSSEM 294 314 POTENTIAL.
FT TRANSSEM 323 343 POTENTIAL.
FT TRANSSEM 347 367 POTENTIAL.
FT TRANSSEM 398 418 POTENTIAL.
FT TRANSSEM 433 453 POTENTIAL.
SQ SEQUENCE 639 AA; 70153 MW; 25DF2819761B415 CRC64;

Query Match 4.9%, Score 135; DB 1; Length 639;
 Best Local Similarity 22.4%; Pred. No. 0.017;
 Matches 104; Conservative 69; Mismatches 161; Indels 130; Gaps 28;

Qy 62 LGLSHALTSMWLGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCMILCVAVIVVGF 121
 Db 59 IGLTALVVIIRL-----AEVIDPILGNIVDNTKRWGKFKPQVIGAVSVLLVVI-F 113
 Qy 122 SSDGAALGDKTKECHSLYHGRWHA-AIVYVGLWLLD-----FSNNTVQGPARAMMADLC 176
 Db 114 TGIFGLA-----HIN-----WTAFAIVFVFLFILLDIFYSADVAYWGVPAISED-- 159
 Qy 177 DHGPSAANSIFCSMMALGNILYSSGSTNNWH-----KWFPELKTSAACEACAN 226
 Db 160 -----SKERGIFTS-----LGSEFTGSI-GWGLTMIIVPVVTVFTFIATGKEHQGSG 206
 Qy 227 LKGAFLVAVFLVCLTVLIFAEVPEYRANENLPTTKAGGEVETPTGPIAVLKGFDL 286
 Db 207 WFGFSIVVSIYAVLSALAVAFGTKE-----KDLIRNAA-----TKKT-----SIKDV 249
 Qy 287 PGMPSVLLVTAITWLSWEPFILDYTDW-----GREIYH-----GDPKGSNAQISAFNEGV 338
 Db 250 FSG-----IHNQDILWIS-LAYLMSLAVVTVNGVLFFYEFKFLVCKP-----NEFW 295
 Qy 339 RVGAFGLLNSVILGFSSFLIEPCMKRYGPRVWVTSNFMVCMVAMAATALISFWSLRDYH 398
 Db 296 IAGAI-----ATVIGFSTAPLPVNLK-----FITRKVLFSIGOMAMIL-----SYL 337
 Qy 399 GYVQDAITANASIKAVCLVLFVAFGLVPLAIVLSPFFAVTAQIATRGGGQGLCTGVNLIS 458
 Db 338 FFIFG-----KTNMMVMTIGILNFETFAQLVVLVSLTDSIEY-----GQ-LKNGRNEA 386
 Qy 459 IVIPOVITIALGAGPW-DALFCKGNIPAF--GVASAFALVGVGVG 499
 Db 387 VV-----LAVRMLDKITG-----AFSNGLVGAITAGMTG 418

RESULT 5
 YD74_SUNY3
 ID YD74_SUNY3 STANDARD; PRT; 544 AA.
 AC P74158;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Hypothetical symporter sl11374.
 GN SL11374.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=11148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97061201; PubMed=8905231;
 RX Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hikosawa M., Suglura M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuoka A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yanada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
 CC (SGF).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D90912; BAA18257.1; -;
 DR InterPro; IPR001927; Na/Gal_symp.
 DR TIGRams; TIGR00792; gph: 1
 DR PROSITE; PS00872; NA_GALACTOSIDE_SYMP; FALSE_NEG
 KW Hypothetical protein; Transport; Transmembrane; Symport;
 KW Complete proteome.
 FT TRANSMEM 31 51 POTENTIAL.
 FT TRANSMEM 52 72 POTENTIAL.
 FT TRANSMEM 84 104 POTENTIAL.
 FT TRANSMEM 116 136 POTENTIAL.
 FT TRANSMEM 162 182 POTENTIAL.
 FT TRANSMEM 191 211 POTENTIAL.
 FT TRANSMEM 230 250 POTENTIAL.
 FT TRANSMEM 257 277 POTENTIAL.
 FT TRANSMEM 318 338 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 383 403 POTENTIAL.
 FT TRANSMEM 407 427 POTENTIAL.
 FT TRANSMEM 450 470 POTENTIAL.
 FT TRANSMEM 501 521 POTENTIAL.
 SQ SEQUENCE 544 AA; 60064 MW; C91D0EDFF32277EE CRC64;

Query Match 4.9%; Score 134; DB 1; Length 544;
 Best Local Similarity 19.3%; Pred. No. 0.017;
 Matches 109; Conservative 77; Mismatches 211; Indels 168; Gaps 25;

Qy 42 GGVOYGWALQLSLTPVY-----QTLGLSHALTSMWLGPIAGLVVQPLVGLYSDRCTS 96
 Db 19 GAGDFGPAITANILVYLLFFLTDVAGVPAALAGSLVMIGKIFDINPITGLISDRTRS 78
 Qy 97 RWGRRRPFILTG-----CMLICVAVIVVGVSSDGAALGDTKEHCSLYHGRWHAIVVVL 152
 Db 79 RWGRRPWLGMIPFALFYTAQWLIHPHSD-----RLTN-QWGLFIYVA 124
 Qy 153 GFWLLDSNNTVQGPARAMMADLCHHGP-SAANSIFCSMMALGNILG-----YSSGST 205
 Db 125 IMAFNLCTYTNVLPYTALTPELTQNYNERTLRNLSRFAFSIGSILSLYLITAAGLP 184
 Qy 206 NNHKKWF-----PFLKTSACCEACANL--KG-----AFLVAVVFL 238
 Db 185 DRPQOFGELGVMISVLSISALLWSALRQKREPILSPSLRRRLAPLMAAGITLILL 244
 Qy 239 VLCTVTLIFAKEVPY-----RANENLPTTKAGGEVETEP--- 273
 Db 245 AIAKSNLGLGGSGEDYISFFLILLGLINGGEGFTLRDSAVEEHLQ-----KLENSPSG 298
 Qy 274 -TGPLAVLKGFDLPFGMPSPVLLVTAITWLSWFP-----FILDYDWMGREIYHGDPK 325
 Db 299 VTENLPKLLQK-LAFSNRAFLFVIGIYLCSWLAVQLTASILVFPVFSWM----- 347
 Qy 326 GSNQAISAFNEGVVRGAFGLLNSVILGFSSFLIEPCMKRYGPRVWVTSNFMVCMVAAA 385
 Db 348 GLNQQS-----GTALAVQGTAL-VMLFVWQALAQFDKKVIYFLGSM---VWMA 395
 Qy 386 TALISFWSLRDHYGVQDAITANASIKAVCLVLFVAFGLVPLAIVLSPFFAVTAQL----- 440
 Db 396 EA--GLWLVQP--GV-----ALLVTLAIFAGVGVSVAYLIPWSMIPDVVDLDE 440
 Qy 441 --AATRGG-----OGLCTGVNLISIVI-----PQVITIALGAGPWDL 476
 Db 441 LNTCKRRGFFYAFPMVLVQLKQGLALGLFLVGLTLEASGFIARIPGEPIPIQPSALSMAIR 500
 Qy 477 FGKGNIPAFGVASAFALVGVGVGF 501
 Db 501 FAVAPLPAP-----FLIGGLILAI 520

RESULT 6
 ID XYL_P_LACPE STANDARD; PRT; 479 AA.
 AC P96792;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)


```
FT TRANSMEM 253 273 POTENTIAL.
FT TRANSMEM 287 307 POTENTIAL.
FT TRANSMEM 308 328 POTENTIAL.
FT TRANSMEM 344 364 POTENTIAL.
FT TRANSMEM 366 386 POTENTIAL.
SQ SEQUENCE 393 AA: 42427 MW; 1447A92FE586F701 CRC64;

Query Match. 4.7%; Score 127; DB 1; Length 393;
Best Local Similarity 20.0%; Pred. No. 0.039;
Matches 94; Conservative 57; Mismatches 151; Indels 168; Gaps 22;

Qy 14 AGVRAAAVHDVAPISLGRLLAGWAGVQYQWALQSLTTPVQT-LGLSHALTFM 72
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 73 WLCGGIAGLVVPLVGLSDRCTSWGRRRPFILTGCMILCVAVIVVGFSSDIGAALGDT 132
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 57 YGSAVIGIVVQILATVSD---RQGRKTLILQCCLLGALA----- 95
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 133 KEHCSLYHGPRHAAIVVYVGLFWLDFSNNTVQGPARAMADLCHDHGFSAA-----N 185
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 96 ---CLLYAARNRYFVLLFI-GVLLSSF-GSTANPOLFALAREHADRTGRGAAMFSSVMRA 150
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 186 SIFCSW-----MALNIIIGYSSGSTNNHKKWPPFLKTSACCEACANLGAFLVAVFLVL 240
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 151 QISLSWVIGPPVAFALALGFG-----FPAM-----YLTAAVVEVL 185
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 241 CLTVTLIFAKEVP-----YRAN 257
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 186 GGLLVLLLPMPKTRVKSATLESPRONRRDTLLLTACTLMTWNCNGIYLINMPYLVLN 245
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 258 E-NLPT-----TKAGGEVETEPTGLAVLKGFKDLPFGMPVSVLLVTATWLSWFP-F 307
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 246 ELRLPEKLAGVMGTAAGLEI-----PVMLLAGVLTSLKRLMLRLAVTAGLIFYTGL 299
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 308 ILYDTPW-----MGREIYHGD---PKGSNAQISAFNEGVVAGFAGLLL 347
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 300 TLLNGSWALLQLLNAIFIGTLACGMV-LYFQDLPQGAAGAAATLFTNTRV---GWII 355
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 348 NSVILGFSSFLIEPCMKRVGRVYV-VTSNEMVVCVAMAATALISFMSLRD 396
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 356 SGSLAGIVA-----EWSYHAGFVIAIAMLAGAACVCMWRIRD 392
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
SL55_RAT STANDARD; PRT; 618 AA.
AC Q63008;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sodium/iodide cotransporter (Na(+)/I(-) cotransporter) (Sodium-iodide
DE symporter) (Na+/I-symporter).
GN SLC5A5 OR NIS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID:10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=96158880; PubMed=8559252;
RA Dal G., Levy O., Carrasco N.;
RT "Cloning and characterization of the thyroid iodide transporter.";
RL Nature 379:458-460(1996).
CC -1- FUNCTION: MEDIATES IODIDE UPTAKE IN THE THYROID GLAND.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SODIUM: Solute symporter family (SSF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U60282; AAB03338.1; -; Na/solut_symport.
DR InterPro: IPR001734; Na/solut_symport.
DR Pfam: PF00474; SSF; 1.
DR TIGRFAMs: TIGR00813; sss; 1.
DR PROSITE: PS00456; NA_SOLUT_SYMP_1; 1.
DR PROSITE: PS00457; NA_SOLUT_SYMP_2; FALSE_NEG.
DR PROSITE: PS0283; NA_SOLUT_SYMP_3; 1.
KW Transport; Transmembrane; Sodium transport; Symport; Glycoprotein;
KW Phosphorylation.
FT DOMAIN 1 16 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 17 37 POTENTIAL.
FT DOMAIN 38 53 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 54 74 POTENTIAL.
FT DOMAIN 75 88 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 89 109 POTENTIAL.
FT DOMAIN 110 136 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 137 157 POTENTIAL.
FT DOMAIN 158 163 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 164 184 POTENTIAL.
FT DOMAIN 185 186 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 187 207 POTENTIAL.
FT DOMAIN 208 241 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 242 262 POTENTIAL.
FT DOMAIN 263 286 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 287 307 POTENTIAL.
FT DOMAIN 308 326 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 327 347 POTENTIAL.
FT DOMAIN 348 391 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 392 412 POTENTIAL.
FT DOMAIN 413 416 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 417 437 POTENTIAL.
FT DOMAIN 438 444 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 445 465 POTENTIAL.
FT DOMAIN 466 520 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 521 541 POTENTIAL.
FT DOMAIN 542 618 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 551 551 PHOSPHORYLATION (BY PKA) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 618 AA; 65196 MW; 91EFAA6752B4F4B2 CRC64;

Query Match 4.6%; Score 126.5; DB 1; Length 618;
Best Local Similarity 21.7%; Pred. No. 0.067;
Matches 121; Conservative 71; Mismatches 210; Indels 155; Gaps 25;

Qy 1 MARG-----DGGQLAELSAAGVGAAGVVDHVPISLGRLLAGWAGVQYQV--- 47
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 36 LARGGORSADDFGTGRLAAVPGVGLSLAASPMASVQ-----VLGVPAEAARYGLKF 87
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 48 W--ALQL--SLLT-----PYVQTGLSHA-----LTSFMWLCGPTAGLVVQPLVGLY 90
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 88 LMMCAQGLNSLLATFLPLFIFRLGLTSTQYLELRFSAVRLCGTLOYLVA----- 140
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 91 SDRCTSRGRRRPFILTGCMILCVAVI---VVGSSDIGAALGDTKEHCSLYHGPRHAA 147
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 141 -----TMLYTGIIVYAPALILNQVTGL--DIWASLLSTGICTLTITVGGMKA 186
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 148 IVY-----VLGFLLDFSNNTVQGPARAMMA-----DLCDHHGSPSAANSIFCS 190
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 187 VVWTDVQVVMVLMGVFVILARGVILLGGRPNVLSLAQNHSRLNLMDFDPPDRSRYTFWT 246
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 191 WMALCNILGYSSGSTNNHKKWPPFLKTSACCEACANLGAFLV--AVFVLCLVTLIF 248
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 247 FIVGTGLVWLSMYGVNQAO-----VQRYVACHTEGKAKLALLNQLGLFLIVASACCGI 301
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 249 AKEVYRANENLPTTKAGGEVETEPTGLAVLKGFKDLPFGMPVSVLLVTATWLSWFPFI 308
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 302 VMFVIYKDCDPLLTGRISA---PDQYMPLLVLDIFEDL-PGVPGFLFACA----- 347
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Query Match 4.6%; Score 126.5; DB 1; Length 618;
Best Local Similarity 21.7%; Pred. No. 0.067;
Matches 121; Conservative 71; Mismatches 210; Indels 155; Gaps 25;
```

30	LPISRLRLSMFG--VSVGMVNVLLVGT-----LNRVWVLEVPASTVIGTMSIPL	80
84	----QPLVGLYSDRCTSRWG-RRRPFLITGCM-----ICVAVIV-GFSDDIGAALG	130
81	IFGKSDTHKSALGWRVRVPYWKGTLLQWGFAIMPEALIVLSQESAAGA---	137
131	DTKEHSLYHGRWHAIVYVVLGFWLLDFSNNTVQGPARAMMADLCDHHGSPSAANSIFCS	190
138	-----PEWIGILSAVSFLVAGVHTVQTVGLALADYDLAPREDQPNVWGLMYV	186
191	WMALGNILYSGSGTNNWIKWFFPKLTSACCEACANLKGAFVAVVFLVCLFTWTLIFAK	250
187	MLLVGMIV--SALLFGWLEDFYHAKLIK-----LOGAAVATMVFNIAL-----WKM	233
251	EV--PYRANENLPTTKAGGVEVETPGPLAVLKGKDLPPGMPSVLLVTAITWLSWFPPI	308
234	EARDVRARQRL-----EGDPEPS-----FRE-----AMGLF--	260
309	LYDTDMGREIYHDPKGSNAQISAFNEGVRVGAFLGILLNSVILGFSSFLIEPMCKRVGP	368
261	-----TRGPNARLLWVIGLTLGFGL-----SDVLEFPFGGVLD	296
369	RVYVNTSNFMVVCVMAATALISF-WSLRDY-HGYVQDAIT---ANASIKAVCLVLFAFLG	423
297	MSVAATTK--LTAAGVAGTLVGAWASRVLSRGYDPMAGWAGVAVGLPAFAITFS---	351
424	VPLAIIYSVPFVATOAATATRGG--OGLCTGVNISIVIPQVITAGAGPDALFGK	480
352	---ATQSEPVFVGTGMMAGFAGLFSHGTLTATMSA---PRAQVGLALGAW-----G	399
481	NIPAFGVASAFALVGVV	497
400	AVQATSAGAGIAL-GGV	415

RESULT 10

YAGG_ECOLI STANDARD; PRT; 460 AA.

AC P75683; P71292;

01-NOV-1997 (Rel. 35, Created)

01-NOV-1997 (Rel. 35, Last sequence update)

16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical symporter yagg.

YAGG OR B0270.

GN Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

[1] SEQUENCE FROM N.A.

RP STRAIN=K12 / MG1655;

RC MEDLINE=97426617; PubMed=9278503;

RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Rileym M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

[2] SEQUENCE FROM N.A.

RP Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,

RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,

RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,

RA Davis R.W.;

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

CC -! SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane

CC (potential).

CC -! SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY

CC (SGF). STRONG, TO E.COLI Y1CJ.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use.


```
QY 362 MCRKVGPRVWVTSNFMVCMVAMATALISFWSLRDYGHVQDAI-----TANAS 410
Db 535 -----TDSLVAIGITAGCVLWGYLLYQGV-DPLGGVKSLLWPLFGISNQM 580

QY 411 IKAVCLVL 418
Db 581 LAVALVL 588

RESULT 12
ID YDK_ECOLI STANDARD; PRT; 459 AA.
AC P76230; P76911;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical metabolite transport protein ydjk.
GN YDK OR B1775.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishio Y., Oshima T., Saito N.,
RA Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 Genome
RT Corresponding to the 28.0-40.1 min Region on the Linkage Map.";
RL DNA Res. 3:363-377(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
FT DOMAIN 112 112 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 113 133 4 (POTENTIAL).
FT DOMAIN 134 153 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 154 174 5 (POTENTIAL).
FT DOMAIN 175 181 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 182 202 6 (POTENTIAL).
FT DOMAIN 203 271 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 272 292 7 (POTENTIAL).
FT DOMAIN 293 301 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 302 322 8 (POTENTIAL).
FT DOMAIN 323 329 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 330 350 9 (POTENTIAL).
FT DOMAIN 351 351 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 352 372 10 (POTENTIAL).
FT DOMAIN 373 399 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 400 420 11 (POTENTIAL).
FT TRANSMEM 421 441 12 (POTENTIAL).
FT DOMAIN 442 459 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 459 AA; 49602 MW; BCAB53ECB8BDD77 CRC64;

Query Match 4.5%; Score 122; DB 1; Length 459;
Best Local Similarity 22.4%; Pred. No. 0.1;
Matches 104; Conservative 54; Mismatches 167; Indels 140; Gaps 25;

QY 132 TKEHCS--LYHGP--RWHAAYVYVGLFWLLDFSNNTVQGPARAMMADL--CDHHGPSAA 185
Db 5 TKPHCGARLDRLPCRWHSSMFAIVAFGLVCWSNAVGLLILAOQLKALGWDNSTTATFS 64

QY 186 SIFCSWML----GNILCYSSGSSNNHKKWPFPLKTSACCACANLKCAFLVAVVFLVLC 241
Db 65 AITTAGMFLGALVGGIIGDKTGRN-----AFILYEA-IHTASVVGAFSPNMDFLIAC 117

QY 242 -----LTVTLIFAKEVPYRANENLPTTKAGGEVETPTGPLA--VLKGFKDLPPG 289
Db 118 RVKMGVGLGALLVL-FAGFTEYMPGRNRTGWSRVSGFNWSYPLCSLIAMGLTPLISA 176

QY 290 -----MPSV--LLVTATITLSWFFPILYDMDMGREIYHGDPKGSKNAQISAFNEG 338
Db 177 EWNVRVQLLIPALLSLIATALAW-RYFP---ESPRWLES---RGRYQAEKVMRSIEGV 229

QY 339 --RVG-----AFGLLLNSVLGPFSSFLIEPWCRCVKPRVV-- 371
Db 230 IROTGKPLPPVVIADGKAPQAVPYSALLTGVLLKRVILG-----SCVLIAANNVVOY 281

QY 372 ---WVTSNFM-----VCVAMAA-----TALISFWSLRDY 397
Db 282 TLINLWPIFTWQGINLKSIVLNTMSMFGAPFGIFIAMLVMDKIPKRTMGVGLLILIAV 341

QY 398 HGYVODAITANASIKAV--CLVLFAPLGVPLAILLYSVP--FAVTAQLAATRGGGGLCTG 453
Db 342 LGYIYSLQTSMLLITLIGFLITFYVMYVCYASAVYVPEIWPTEAKLR-----GSGLANA 396

QY 454 VLNIS-IVIPQVIITAGAGPDALFGKGNIPAFGVASAFALVGGV 497
Db 397 VGRISGIAAPYAVAVL-----LSSYGVTVGVFILLGAV 428

RESULT 13
YICJ_ECOLI STANDARD; PRT; 460 AA.
AC P31435; P76724;
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical symporter yicJ.
GN YICJ OR B3657.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
```



```

CC -----
DR EMBL; D90203; BAA14228.1;
DR EMBL; AE008727; AAL19633.1;
DR EMBL; AL627267; CAD05152.1;
DR PIR; JQ0576; JQ0576
DR StyGene; SG10058; cita.
DR InterPro; IPR004736; Cit_H_sympor.
DR Pfam; PF00083; sugar tr; 1.
DR TIGRfams; TIGR00883; 2A0106; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Citrate utilization; Transport; Transmembrane; Inner membrane;
KW Symport; Complete proteome.
FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 22 42 1 (POTENTIAL).
FT DOMAIN 43 54 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 55 75 2 (POTENTIAL).
FT DOMAIN 76 87 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 88 108 3 (POTENTIAL).
FT DOMAIN 109 111 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 112 132 4 (POTENTIAL).
FT DOMAIN 133 164 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 165 185 5 (POTENTIAL).
FT DOMAIN 186 186 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 187 207 6 (POTENTIAL).
FT DOMAIN 208 238 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 239 259 7 (POTENTIAL).
FT DOMAIN 260 276 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 277 297 8 (POTENTIAL).
FT DOMAIN 298 304 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 305 325 9 (POTENTIAL).
FT DOMAIN 326 335 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 336 356 10 (POTENTIAL).
FT DOMAIN 357 366 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 367 387 11 (POTENTIAL).
FT DOMAIN 388 400 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 401 421 12 (POTENTIAL).
FT DOMAIN 422 434 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 434 AA; 47188 MW; E0F077EAC70DB444 CRC64;

Query Match 4.48; Score 120; DB 1; Length 434;
Best Local Similarity 20.28; Pred. No. 0.14;
Matches 98; Conservative 73; Mismatches 165; Indels 148; Gaps 24;

QY 61 TLGLSHALTSPWMLCGPIAGLVQPLVGLYSDRCTSRGRRRPFLIT----GCMLICVAV 116
DB 54 SLMLTFVFGSGFLMRPVGAIV---LQAYIDRI---GRKGLMVTLAIMGCGTLLIA- 104
QY 117 IVVGFSSDIGAALGDKHECHSLYHGPRWHAIVVYLVFWLDFSNNTVQGPAMMADLC 176
DB 105 LVPGYQT-IGLA-----APALVLLGRLLGQFSAGVGLGVSVVLSEI- 145
QY 177 DHHGPSAANSIFCSW-----MALGNILGYSGST-----NNWHKWFPLKTSACCE 222
DB 146 ---ATPGKNGFYTSQMSAQVAIVVAALIGYSLNITLGHDAISEW-GW----- 190
QY 223 ACANLKAFLVAVVFLVLCVTLTFLFAKEVYRANENLPTTKAGEVEVETPTGP---LAV 279
DB 191 -----RIPFFICMIIPLFV-----LRSQTEAFLOKRPDTRFIFATI 233
QY 280 LKGFKDLPPGMPSVLLVTAITWLSWFFPLYDMDMGREIYHGPDKGNAOISAFNEGVR 339
DB 234 AKNRRIITG---TLLVAMTV--TFYFYTIVTYTGRV-----LN 270
QY 340 VCAFGLLNSVLGSSFLIEPMCKVKVPRVWVTSNFWCVMAATALISFWSLRDHYG 399
DB 271 LSARDSLVTMLGVGSNFIWPIPIGGAISDRI-----GRRVLMGTLLALITWPMQW-- 324
QY 400 YVQDAITANASIKAVCLVL--FAFL-----GVPLAIL-----YSVPFAVTAQLAATR 444
DB 325 -----LTRAPEFTRMTLVLLWFSFFGNGYNGAMVAALTEVMPVVRVTVGFSLSLATAI 379

```

```

QY 445 GGGGLCTGVNLNISIVIPQVILIALGAGPMDALFCGNIPAFGVASAFALGVGVGVFLFP 504
DB 380 FGG-----LTPAISTAL-----VKLTGDKSSPGWLMCA-ALCGLAATAMLFV 421
QY 505 KISK 508
DB 422 RLSR 425
RESULT 15
ARAE_BACSU
ID ARAE_BACSU STANDARD; PRT; 464 AA.
AC P96710;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Arabinose-proton symporter (Arabinose transporter).
GN ARAE.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
subtilis."
RL Nature 390:249-256(1997).
RN [2]
RP SEQUENCE OF 1-223 FROM N.A.
RC STRAIN=168;
RX MEDLINE=97197523; PubMed=9045819;
RA Sa-Nogueira S., Mota L.J.;
RT "Negative regulation of L-arabinose metabolism in Bacillus subtilis:
characterization of the arar (araC) gene."
RN [3]
RP SEQUENCE OF 223-464 FROM N.A., AND FUNCTION.
RC STRAIN=168;
RX MEDLINE=98062200; PubMed=9401028;
RA Sa-Nogueira I., Ramos S.S.;
RT "Cloning, functional analysis, and transcriptional regulation of the
Bacillus subtilis arae gene involved in L-arabinose utilization."
RN [4]
RP J. Bacteriol. 179:7705-7711(1997).
CC -1- FUNCTION: UPTAKE OF ARABINOSE ACROSS THE BOUNDARY MEMBRANE WITH

```

CC THE CONCOMITANT EXPORT OF A PROTON (SYMPTOM SYSTEM).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z99121; CAB15401.1; -
 DR EMBL: X98354; CAA66998.1; -
 DR EMBL: Y12105; CAA72812.1; -
 DR Subtilist: BG11907; araE.
 DR InterPro: IPR003663; CHO_transport.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 DR PRINTS: PR00171; SUGRTNSPORT.
 DR TIGRFAMS: TIGR00879; SP; 1.
 DR PROSITE: PS00216; SUGAR_TRANSPORT_1; 2.
 DR PROSITE: PS00217; SUGAR_TRANSPORT_2; 1.
 KW Transport; Sugar transport; Transmembrane; Symport; Complete proteome.
 FT TRANSMEM 21 43 POTENTIAL.
 FT TRANSMEM 63 85 POTENTIAL.
 FT TRANSMEM 92 111 POTENTIAL.
 FT TRANSMEM 116 138 POTENTIAL.
 FT TRANSMEM 150 172 POTENTIAL.
 FT TRANSMEM 185 207 POTENTIAL.
 FT TRANSMEM 266 288 POTENTIAL.
 FT TRANSMEM 303 325 POTENTIAL.
 FT TRANSMEM 332 354 POTENTIAL.
 FT TRANSMEM 364 386 POTENTIAL.
 FT TRANSMEM 398 420 POTENTIAL.
 FT TRANSMEM 424 446 POTENTIAL.
 SQ SEQUENCE 464 AA; 50411 MW; 13B417061CB61DA2 CRC64;
 Query Match 4.4%; Score 119.5; DB 1; Length 464;
 Best Local Similarity 21.3%; Pred. No. 0.16;
 Matches 111; Conservative 67; Mismatches 174; Indels 169; Gaps 27;
 QY 30 SLGRLLIAGVA--GGVOYGV-----ALQLSLTPYVOTGLSHALTSFWMWLG 76
 DB 19 SMGFVILISCAAGGLGGLLYDYTAIVSAIGCFKLDLYSLSPFMEGLVISSIM----- 70
 QY 77 PIAGLVQPLVGLYSDRCTRWGRRRPFILTCMLICVAVIVGVFSSDIGAALGDTKEHC 136
 DB 71 -IGVGVGISGFLSD-----RFGRRK-ILMTAALLFAISAIVSALSQDVSTLI----- 117
 QY 137 SLVHGRPRHAAIVVVLGFWLLDFSNNTVOGPARAMADLCDHHGSPSAANSIFCSWMAALGN 196
 DB 118 -----IARIIGGLI-----GMGSLSVTYITEAAPPAAIRGSLSLYQLFT 158
 QY 197 ILGYSS-----GSTNNW-----HKWPFPLKTSACCEACANLGAFLVAVVFLVLC 241
 DB 159 ILGISATYFINLAVQSGTVEGWVHTGWRWN-----LAYGMVPSVIFPLVL 204
 QY 242 LVTYLIFAKEVPY-----RANENLP-TTKAGG-----VETPTGPLAVLK 281
 DB 205 LVV-----PSPRLAKAGKTNEALKTLTRINGETVAKEELKNIENTSKIEQMGSLSQL- 258
 QY 282 GFKDLPQMPSSLVLTATITLSNFPFFLYDTDMGRIYHGDPKGSNAQISAFNEGRVVG 341
 DB 259 -FK---PGLRKALVIGILLALFNQVIGMNAITYYGPIFKMMFGQNA---GFVTTICVG 311
 QY 342 AFGLLNSVILGFSFLIEPMCRKVGPRVVMVTSNFM--VCVMAATATLISFWSLRDHYGY 400
 DB 312 VVEVIEFVIAV---LLIDKVRK---KLMSIGSAFAFIEMILIGT---SFY----- 353
 QY 401 VQDAITANASIKAVCLVFLFPLGVLPAIILYSVPFAVTAQAANRGGGGGLCTGVNLNLSIV 460
 DB 354 ----FELTSGMTIMIVLI-----LGFVAACVSV-----GPITWIM-ISEI 388

QY 461 IPOVVIAGAG-----PWDALFGKGN-----IPAFGVASAF 491
 DB 389 FPNHLRARAGIATIFLWGANWAIGQFVPMWIDSGLAYTF 429
 RESULT 16
 YICM_ECOLI
 ID YICM_ECOLI STANDARD; PRT; 412 AA.
 AC P31438; P76725;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 30, Last sequence update)
 DE Hypothetical protein yicm.
 GN YICM OR B3662.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=K12 / MG1655;
 RX MEDLINE=93315143; PubMed=7686882;
 RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
 RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
 RL genome: organizational symmetry around the origin of replication.";
 RL Genomics 16:551-561(1993).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L10328; AAA62014.1; ALT_INIT.
 DR EMBL: AE000444; AAC76685.1; ALT_INIT.
 DR EcoGene; EG11689; yicm.
 DR InterPro: IPR003662; sub_transporter.
 DR Pfam: PF00083; sugar_tr; 1.
 KW Hypothetical protein; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 102 122 POTENTIAL.
 FT TRANSMEM 124 144 POTENTIAL.
 FT TRANSMEM 164 184 POTENTIAL.
 FT TRANSMEM 192 212 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 FT TRANSMEM 298 318 POTENTIAL.
 FT TRANSMEM 322 342 POTENTIAL.
 FT TRANSMEM 360 380 POTENTIAL.
 FT TRANSMEM 383 403 POTENTIAL.
 SQ SEQUENCE 412 AA; 43626 MW; 280CFD30C6912B2 CRC64;
 Query Match 4.4%; Score 119; DB 1; Length 412;
 Best Local Similarity 20.1%; Pred. No. 0.15;
 Matches 107; Conservative 76; Mismatches 155; Indels 194; Gaps 28;
 QY 6 GGQALAEISAGVGAADV--DHVAPISLGRLLIAGMAGVGVQGNALQSLTLLPVVOTLG 63
 DB 14 GNTMSEFIAENRGADAITRPNWSAVFVAFVACVLI---VEF---LPVSLTTPMAODLG 67
 QY 64 LSHALTSFWMWLCGPIAG--LVVQPLVGLYSD---RCTSRWGRRRPFILTCMLICVAVIV 118
 DB 68 ISEG-----VAGQSVTVTAFAVAFSLFTQTQTQATDRYVIVFLVLTLSCLL 117
 QY 119 VGFSDD-----IGAALGDTKEHCSLYHGRPHAAIVVVLGFWLLDFSNNTVOGPA 168
 DB 119 VGFSDD-----IGAALGDTKEHCSLYHGRPHAAIVVVLGFWLLDFSNNTVOGPA 168

Db 118 VSFANSFSLLLIGRACLGALG-----GFWMSASLTMLRVLPP 155
Qy 169 RAMMADLCHDHGSPSANSIFCS-----WMALGNILYSSGSTNNHKKWFFFLKTSACC 221
Db 156 RIV-----PKALSIVFGAVSTALVIAAPLGSFLGELIG-----W----- 189
Qy 222 ECANKLGAFLVAVVFLVCLVTLIFAKEVPRANENLPTTKAGGEVETETPTGPLAVLK 281
Db 190 -----RNVFNAAVMGVLCI--EWIILSLP-----SLP-----GEPHOKONTFRLLQ 230
Qy 282 GFKDLPKGPMSVLLVTAITWLSWFFILYDTDMGREIYHGDPKGNSAQISAFNEGVAVG 341
Db 231 -----RPGVMAGMIAIFMSFAGQFAFFTYI-----RPVY-----MNLAF-----G 266
Qy 342 AFGLLNSVLGFSFLIEPMCKRKGPRVWVTSNFM-----VGVMAATALISFWSLRDY 397
Db 267 VDLGLVLLSFGIASF-----IGTSL-----SFLKRSVKLALACAPLI----- 306
Qy 398 HGVYODAITANASIKAVCLVLF-----FLGVPLAILYSVPEA-----VTAQLA--AT 443
Db 307 -----LAVSALVLTLMGSDKIVATGVAILIWLGLTFALVPVGVSTWITRSLADQAE 355
Qy 444 RGGGGLCTGVNLISIVIPQVILALGAPWDALFGKNIPARFVASAFALVG 495
Db 356 KAG-----SIQVAVIQLANTCCAA-----IGGYALDNIGLTSPLMLSG 393

RESULT 17
MOT4_CHICK
ID MOT4_CHICK STANDARD; PRT; 473 AA.
AC P57788; Q9DEY5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Monocarboxylate transporter 4 (MCT 4).
GN SLC16A3 OR MCT4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoon H., Philip N.
RT "Characterization of the chicken and mouse MCT4 genes: tissue
distribution and relationship to other monocarboxylate transporters.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROTON-LINKED MONOCARBOXYLATE TRANSPORTER. CATALYZES THE
RAPID TRANSPORT ACROSS THE PLASMA MEMBRANE OF MANY
MONOCARBOXYLATES SUCH AS LACTATE, PYRUVATE, BRANCHED-CHAIN OXO
ACIDS DERIVED FROM LEUCINE, VALINE AND ISOLEUCINE, AND THE KETONE
BODIES ACETOACETATE, BETA-HYDROXYBUTYRATE AND ACETATE (BY
SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Plasma membrane
(By similarity).
CC -!- SIMILARITY: BELONGS TO THE SLC16 FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AF204396; AAF67524.1; -
DR EMBL; AF308452; AAG25703.1; -
DR InterPro; IPR004743; Mcarb.transprot.
DR TIGRFAMs; TIGR00892; 2A0113.1.
KW Transport; Symport; Transmembrane; Multigene family.
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 18 38 POTENTIAL.
FT DOMAIN 39 61 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 62 82
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 83 91
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 92 112
EXTRACELLULAR (POTENTIAL).
FT DOMAIN 113 115
EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 116 136
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 137 149
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 150 170
EXTRACELLULAR (POTENTIAL).
FT DOMAIN 171 179
EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 180 200
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 201 231
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 232 252
EXTRACELLULAR (POTENTIAL).
FT DOMAIN 253 268
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 269 289
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 290 297
EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 298 318
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 319 321
EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 322 342
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 343 358
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 359 379
EXTRACELLULAR (POTENTIAL).
FT DOMAIN 380 388
CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 389 409
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 410 473
CYTOPLASMIC (POTENTIAL).
FT CONFLICT 103 103
A -> V (IN REF. 1; AAG25703).
SQ SEQUENCE 473 AA; 51041 MW; 87B8F98E675CC7E8 CRC64;

Query Match 4.3%; Score 118.5; DB 1; Length 473;
Best Local Similarity 20.7%; Pred. No. 0.19;
Matches 93; Conservative 64; Mismatches 173; Indels 120; Gaps 24;

Qy 18 GAAAVVDHVPISL-----GRILLAG-MVAGGVQVGMALQISLL-----TPYVQTL 62
Db 2 GAVVDDGPGVKAPDGGGWNVLFGCFITGFSYAFFKAVSVFFKELIREFGVGSYDTA 61
Qy 63 GLSHALTSFWMLCGPIAGLVQPLVGLYSDRCTSEWRRRPFILTCMLICVAVIVVGS 122
Db 62 WISSILLAMLYGTGPLCSV-----CVNRFQ-CRPVNLVGGFLFASGMMAISFC 108
Qy 123 SDIGAALGDTREHSCSYHGRWHAIAVYVLGFWLDFSNNTVQGPARAMMADLCHDHGP- 181
Db 109 TSI-----VQIY-----LTAGVITGLG-ALNFQ-----PSLIMLNRYFDKRRL 147
Qy 182 ----SAANS--IFCSWMALGNILYSSGSTNNHKKWFFFLKTSACCACANLKGAFLVAV 235
Db 148 ANGLSAGSPVFLCALSPGLQILOHEY-----W-----RGFLILG 184
Qy 236 VFLVCLVTLIFAKEVPRANENLPTTKAGGEVETETPTGPLAVLKGFKDLPPGMPVLL 295
Db 185 GMLNCCVCGALMRPLEPPKSE---ATREPAEKAKK--KLDDPSVFKD--GGVVIYTL 237
Qy 296 VTATWLSWFFILYDTDMGREIYHGDPKGNSAQISAFNEGVAVGAFGLLNSVLGFS 355
Db 238 AASIMVLGLFVPPVSVY-AKDLGYQDTK-----AFL-----TILGFI 277
Qy 356 SFLIEPMCKRKGPRVWVTSNFMVCMVMAATALISFWSLRDYHGVYQDQAITANASIKAVC 415
Db 278 DIFARPICGMVA-GLKWRPR---CVYLFSAFI-----FNGFTD--LMGSMSSVDYGG 324
Qy 416 LVFLA-FLGVPLAILYSVPEAVTAOLAATR 444
Db 325 LVVCFIFGISYGMVGAQFVFLMAIVGTQ 354

RESULT 18
SL55_HUMAN
ID SL55_HUMAN STANDARD; PRT; 643 AA.
AC Q92911; Q43702; Q9NYB6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sodium/iodide cotransporter (Na(+)/I(-) cotransporter) (Sodium-iodide
symporter) (Na+/I-symporter).
GN SLC5A5 OR NIS.
OS Homo sapiens (Human).


```
FT VARIANT 556 556 /FTID=VAR_010269.
FT S -> Q. /FTID=VAR_010270.
SQ SEQUENCE 643 AA: 68666 MW: 02D361A27B2FDA43 CRC64;

Query Match
Best Local Similarity 4.3%; Score 118.5; DB 1; Length 643;
Matches 118; Conservative 72; Mismatches 208; Indels 179; Gaps 25;

QY 29 ISLGRLLILAGVAGG-----VOYQWALQSLTLPYVOTLGL-SHAL--- 68
DB 26 VTFGLGLWGLARGQORNAEDFTGCRRLAALPVLGLSLASFMSA-VQVLGVPSAYRG 84
QY 69 TSFMMLC-----GPIAGLWQPL---VGLXS--DRCTSRMGR-----RPFIL 106
DB 85 LKFLWMLCLGQLNSVLTALLFMPFVYRLGLTSTYVLEMFESRAVRVLCGTLQYIVATMLY 144
QY 107 TCMCLICAVI---VVGSSDGAALGDTKEHCSYHGPVHAAIVY-----VLGF 154
DB 145 TGIVYAPALILNQVTGL--DIWASLLSTGIITCTYTAVGGMKAVVWTDVFOVVMLSGF 202
QY 155 WLDFESNNTVQGPARAMMA-----DLCDHGSPSAANSIFCSNMALGNILGYSSGSTN 206
DB 203 WVLARGVMLVGGPROVLTLAGNHSRINLMDPNPDRSRYTFWTFVVGGLTWLWLSMYGVN 262
QY 207 -----NWHKWFPELTKTSACCACANLKGAFILVAVVFLVCLTV 244
DB 263 QAQVRYVACRPEKOKALLINQVGLFLIVSSAACCG-----IVMFVYTDGCP 313
QY 245 TLIFAKEVPIYRANENLPTTKAGGEVET-EPGCPVLAVLGFKDLPGMPSVLLIVTAITWLS 303
DB 314 LL-----GRISAPDQYMLPLVLDIFEDL-PCVPGLFLACA----- 347
QY 304 WPFILYDIDWGREIYHGDPRKGSNAQISAFNEGVVGVGAFGLLNSVILGSSFLIEPMC 363
DB 348 -----YSGTLSTASTSINAM-AAVTV-----EDLIKRL 375
QY 364 RKVGRVWVWTSNFMVCMVAMAATALISFWSLRDHYGVQDAITANASIKAVCLVLFALF 423
DB 376 RSLAPRKVLIIKGLSLIYGSACLTVAALSLLGGVGLQSGFTVMGVISGPLLGAFL-ILG 434
QY 424 VPLAILYSVPFAVTAQLAATRGGGGGLCTGVNLINISIVIPQVILALGAGPWALFGKNIP 483
DB 435 MFL-----PACNTPGVLAGLGAAL-----SLWVALGATLYPPSEOTMRVL 476
QY 484 AFGVASAFAL---VGVGVGVFLLPKISKROFRAVSAG 517
DB 477 PSSAARCVALSVNASGLDLPALLP---ANDSSRAPSSG 511

RESULT 19
Y4XM_RHISN
ID Y4XM_RHISN STANDARD; PRT; 404 AA.
AC P55705;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical transport protein Y4XM.
GN Y4XM.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(Potential).
```

DT 01-NOV-1997 (Rel. 35, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative 3-hydroxyphenylpropionic acid transporter.
 GN MHPT OR 80353.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=37426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RA "The complete genome sequence of Escherichia coli K-12";
 RT Science 277:1453-1474(1997).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
 RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,
 RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA Nishimoto H., Saito N.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 83-330 FROM N.A.
 RC STRAIN-K12 / CS520;
 RA Fernandez A., Garcia J.L., Diaz E.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD BE A TRANSPORTER FOR 3-PHENYLPROPIONATE
 CC (HYDROCINNAMIC ACID).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (Potential).
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
 CC FRAMESHIFTS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AE000142; AAC73456.1; ALT INIT.
 DR EMBL; U73857; AAB18077.1; ALT INIT.
 DR EMBL; D85613; -; NOT_ANNOTATED_CDS.
 DR EMBL; X97543; CAA66145.1; -;
 DR Ecogene; EGI3293; mhpt.
 DR InterPro; IPR004746; Bnzte transporter.
 DR InterPro; IPR003662; sub transporter.
 DR Pfam; PF00083; sugar tr. 1.
 DR TIGRFAMs; TIGR00895; 2A0115; 1.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
 DR Transport; Transmembrane; Inner membrane; Symport; Complete proteome.
 KW DOMAIN 1 16
 KW TRANSSEM 17 37
 KW DOMAIN 38 53
 KW TRANSSEM 54 74
 KW DOMAIN 75 81
 KW TRANSSEM 82 102
 KW DOMAIN 103 105
 KW TRANSSEM 106 126
 KW DOMAIN 127 142
 KW TRANSSEM 143 163
 KW TRANSSEM 143 163

FT DOMAIN 164 164 PERIPLASMIC (POTENTIAL).
 FT TRANSSEM 165 185 6 (POTENTIAL).
 FT DOMAIN 186 217 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 218 238 7 (POTENTIAL).
 FT DOMAIN 239 253 PERIPLASMIC (POTENTIAL).
 FT TRANSSEM 254 274 8 (POTENTIAL).
 FT DOMAIN 275 279 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 280 300 9 (POTENTIAL).
 FT DOMAIN 301 306 PERIPLASMIC (POTENTIAL).
 FT TRANSSEM 307 327 10 (POTENTIAL).
 FT DOMAIN 328 339 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 340 360 11 (POTENTIAL).
 FT DOMAIN 361 369 PERIPLASMIC (POTENTIAL).
 FT TRANSSEM 370 390 12 (POTENTIAL).
 FT DOMAIN 391 403 CYTOPLASMIC (POTENTIAL).
 FT CONFLICT 295 295 * L-> V (IN REF. 4).
 SQ SEQUENCE * 403 AA; 41550 MW; 85EE921E57686FEB CRC64;
 Query Match 4.3%; Score 117.5; DB 1; Length 403;
 Best Local Similarity 19.6%; Pred. No. 0.19;
 Matches 103; Conservative
 Qy 37 AGYVAGV-----QYGNALQLSLTPVQVTLGLSHALTSMWLCGPIAGLVVQPLV 87
 Db 34 AGIAGGIAQAQAFALDKMQMGWIFSAGI-----LGL-----LPGALVG----- 70
 Qy 88 GLYSDRCTSRWRRPEI---LTCMLICVAV-----IVVGFSSDIGAALGDT 132
 Db 71 GMLAD-----RYGKRILIGVALFLSLATAIAMDWPSLVFARLMTGVG--LGAALPNL 124
 Qy 133 KECHSLYHGRHAAIYVVLGVFLWLLDFSNNVTQGPARAMADLCHDHGPGSAANSTFC 190
 Db 125 IALTSEAGPRERG-----TAVSLMYCGVP 149
 Qy 191 -WMALGNILGYSSGSTNNHKKWF-----PFLKTSACCEACANLKGALVAVVFLVCLT 243
 Db 150 IGAALAAATLGF-AGANLAWQTVFWVGVVP-----LILVPLLMRWLP 190
 Qy 244 VTLIFAEVPRANENLPTTKAGEVEVEPTGPLAVLKGKDLPPGMPFVLLVTAITWL- 302
 Db 191 ESAVFA-----GEKQAP--PLRALFA-----PETATATLLLMWC 223
 Qy 303 -----SWFPFTLYDTDMGREIYHGDPKGSNAQISAFNEGVYRVGAFLLLNSV 350
 Db 224 YFTLLVYVYMLINMLPLLVEQF-----QPSQAAGVVMFALQMGAAAG-- 266
 Qy 351 ILGSSFLIEPCMKRVGRVWVTSNFMVYVMAATAALISFWSLDYHGYVDATANAS 410
 Db 267 -----TLMGLAMDKLRP-----VTMSLLIYSGMLA-SULALGTVSSFNGLIAGFVAG-- 314
 Qy 411 IKAVCLVLFAFLGVPLATLYSV-PFAVTAQLAATRGSGGLCTGVNLN--STIVPQVIA 467
 Db 315 -----LFATGG--QSVLYALAPLFYSQIRAT-GVCTAVAVGRIGAMSGPLAGKMLA 364
 Qy 468 LGAGPMDALFGKGNIPAFGVSASFA---LVGGVGVFLLPKIKSRKQ 510
 Db 365 LGTG-----TVGVMAASAPGILVAGLAVFILMRSRRIQ 398
 RESULT 21
 YPUM_RHOCA
 ID YPUM_RHOCA STANDARD; PRT; 477 AA.
 AC P26176;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 50.4 kDa protein in PUHA-BCHM intergenic region (ORF477)
 DE (Protein F1696).
 OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 OC Rhodobacter.
 OX NCBI_TaxID=1061;
 RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=84259352; PubMed=6744416;
RA Youvan D.C., Bylina E.J., Alberti M., Begusch H., Hearst J.E.;
RT "Nucleotide and deduced polypeptide sequences of the photosynthetic
RT reaction-center, B870 antenna, and flanking polypeptides from R.
RL capsulata."; 1984).
RC Cell 37:949-957(1984).
CC -!- SIMILARITY: TO RHODOBACTER CAPSULATUS PUCC AND TO
CC BACTERIOCHLOROPHYLL SYNTHASE 44.5 KDA CHAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z11165; CAA77521.1; -
DR EMBL; K01183; -; NOT_ANNOTATED_CDS.
DR PIR; B28988; B28988.
DR InterPro; IPR004896; PUCC.
DR Pfam; PF03209; PUCC; 1.
KW Photosynthesis; Hypothetical protein.
SQ SEQUENCE 477 AA; 50372 MW; 09A6A50AECDD4D998 CRC64;

Query Match 4.3%; Score 116.5; DB 1; Length 477;
Best Local Similarity 21.1%; Pred. No. 0.27;
Matches 110; Conservative 73; Mismatches 164; Indels 175; Gaps 29;

QY 34 LILAGWAGGVGYGM-----ALQSLPPYVQTLGLSHALSFEMWLCGPAGL 81
DB 1 MTLRRMIGSLAMTWLPFADAASETLPLRLRLSLFQVSGMAQVL-----LLGTNRV 55
QY 82 VYQVLGVLYSDRCTSRWRRRRFILTGMCLICQVAVIVGVFSSDIG-----AALGDYKE 134
DB 56 MILEL-GV-----PALVVAAM-ISPVLVAPFRAILGHRSDTVRSALG-WKR 99
QY 135 HCSLHGPRHAAIVVYLFGLWLLDFSNITVQGPARAMADLCHRGPSAANSIFCSWMA 194
DB 100 VPYLFGLSLWQGLALMPFSLILSGDQTMGPA-----WAG- 136
QY 195 GNILGYSSGSTNNHKKWPFPLKTSACCENAKLGAFLVAVVFLVLCVTLVLIFAKEVPY 254
DB 137 -----EAFAGV-AFLMAGVGMHMTQTAGLAAAD--- 164
QY 255 RANENLPTTKAGGEVETP-----TGPLAVLKGFKDLPFGMPVLLVLTATW 301
DB 165 RATE-----ETRPQVALLYVFLGMGISAIVGW--LIRDQDTLIRVVG 211
QY 302 LSWFFPFIYDWMGREIYHGDPKGS-NAQISAFNEGVVRVGAFL-----LLNSVI 351
DB 212 CGAMTLVLNVIALWQEVNRPMTKAERAPROSFPRE-----AWGLLAETGALRLATVM 266
QY 352 LGFSSF-----LIEP-----MCRKYGPRVWVTSNFMVCVMAATALISF-WSLRDYHYG 400
DB 267 VGTFLAFSMODVLEPYGGQVLKVG-QTTWLTAGW-----AFGALVGFIVSARRLS- 317
QY 401 VQDAITANASIKAVCLVLFALVGLVPLAIIYVFPFAVTAQLAATRGGGGGLCTGVLLNI-SI 459
DB 318 -QGAHVRVAARGLLVGVAF-----TAVLFS-PLFGSKVLFFASAMGIGLGGMGFIATL 371
QY 460 VTPQVITIALGA-----GPWDALFGKGNIPAFGVASAFALVGG 496
DB 372 TVAMVVVVRGSGIALGAW-----GAAQATAGLA-----VFIGG 406

RESULT 22
ID STPL_ATHATH STANDARD; PRT; 522 AA.
AC P23586; Q9SXB1;
DT 01-NOV-1991 (Rel. 20, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

15-JUN-2002 (Rel. 41, Last annotation update)
DE Glucose transporter (Sugar carrier).
GN STPL OR A1G11260 OR T28P6.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=91005995; PubMed=2209537;
RA Sauer N., Friedlaender K., Graeml-Wicke U.;
RT "Primary structure, genomic organization and heterologous expression
RT of a glucose transporter from Arabidopsis thaliana.";
RL EMBO J. 9:3045-3050(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RA SSP consortium (Salk/Stanford/PGEC).";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ACTIVE UPTAKE OF HEXOSES. PROBABLE GLUCOSE/HYDROGEN
CC SYMPORT.
CC
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X53350; CAA3037.1; -
DR EMBL; AC007259; AAB49995.1; -
DR EMBL; AY059781; AAL24129.1; -
DR EMBL; AY054249; AAL06908.1; -
DR PIR; S12042; S12042.
DR InterPro; IPR003663; CHO_transport.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PRINTS; PR00171; SUGTRNSPORT.
DR TIGRFAMs; TIGR00879; SP; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
KW Transmembrane; transport; Sugar transporter; Symport.
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DOMAIN 1 23 CYTOPLASMIC (POTENTIAL).

```
FT TRANSMEM 24 44 1 (POTENTIAL).
FT TRANSMEM 84 104 2 (POTENTIAL).
FT TRANSMEM 118 138 3 (POTENTIAL).
FT TRANSMEM 141 161 4 (POTENTIAL).
FT TRANSMEM 170 190 5 (POTENTIAL).
FT TRANSMEM 203 223 6 (POTENTIAL).
FT TRANSMEM 296 318 7 (POTENTIAL).
FT TRANSMEM 325 345 8 (POTENTIAL).
FT TRANSMEM 349 369 9 (POTENTIAL).
FT TRANSMEM 385 405 10 (POTENTIAL).
FT TRANSMEM 431 451 11 (POTENTIAL).
FT TRANSMEM 454 474 12 (POTENTIAL).
FT DOMAIN 475 522 CYTOPLASMIC (POTENTIAL).
FT CONFLICT 333 333 A -> G (IN REF. 1).
SQ SEQUENCE 522 AA; 57610 MW; 813AFC8AFA7AA583 CRC64;

Query Match 4.38; Score 116.5; DB 1; Length 522;
Best Local Similarity 23.8%; Pred. No. 0.29;
Matches 70; Conservative 34; Mismatches 85; Indels 105; Gaps 17;

Qy 16 VRGAAVVDHVPISLGLILAGVAGVQVQVGA-----LQLSLTPYVQ----- 60
Db 245 IRG-----VDDVSO-EFDDLVAAKESQSIHPWNLRLRRKYPHPTWAVMIPFFQOLTGI 299

Qy 61 -----TGLSHALTSMWLCGPIAGL--VVOPLVGLYSDRCTSRWRRRPPFIL 106
Db 300 NVIMFYAPVLENTIGFT---TDSLMSAVVTGSVNVAAATLSIYG---VDRWGRRLFLLE 353

Qy 107 TGC-MLICVAVIVVGFSSDIAAGLDYKHCSLYHGRPHAAIV-----YVLGF----- 154
Db 354 GGTOMLICQAVV-----AACIGAKFGVDGTPGEL---PKWAIYVVVTFICIYVAGFAWSWG 406

Qy 155 ---WLLDPSNNTVQGPARAMADLCHRHGPSAANSIFCSWALGNILYSGSGSTNNHKKW 211
Db 407 PLGNLV-----PSEIFPLEI-----RSAQSI-----TVSVNMI 435

Qy 212 PFPLKTSACCBACANLK-GAFVAVVFLVLCVTLVTLIFAKEVPYRANENLPTTK 264
Db 436 PTFIIAQIFLTMCHLKFGLFVFAFFVVMVSIFVIF-----LPETK 478

RESULT 23
ID YNAJ_BACSU STANDARD; PRT; 463 AA.
AC P94488;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical symporter ynaJ.
GN YNAJ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Borchert S., Klein C., Piksa B., Hammelmann M., Entian K.D.;
RT "Sequencing of a 26 kb region of the Bacillus subtilis genome downstream of spoJ."
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchsi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
```

```
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Potwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weizenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC (SGF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U66480; ABA41090.1; -
CC EMBL; Z99113; CAB13641.1; -
CC Subtilist; BG12262; ynaJ.
CC InterPro; IPR001927; Na/Gal_symp.
CC TIGRFAMS; TIGR00792; gph; 1.
CC PROSITE; PS00872; NA_GALACTOSIDE_SYMP; 1.
KW Hypothetical protein; Transport; Transmembrane; Symport;
KW Complete proteome.
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 271 291 POTENTIAL.
FT TRANSMEM 311 331 POTENTIAL.
FT TRANSMEM 334 354 POTENTIAL.
FT TRANSMEM 367 387 POTENTIAL.
FT TRANSMEM 408 428 POTENTIAL.
SQ SEQUENCE 463 AA; 51534 MW; B5F0733048023FF2 CRC64;

Query Match 4.2%; Score 115; DB 1; Length 463;
Best Local Similarity 19.9%; Pred. No. 0.33;
Matches 95; Conservative 63; Mismatches 171; Indels 148; Gaps 20;

Qy 54 LLTFYVQTLGLSHALTSEFMWLCGPIAGLVQPLVGLYSDRCTSRWRRRPPILTGC---M 110
Db 35 LLFFYTDVFGLSAAAGTMFLVRIIADALDPFGITIVDRNTRFARFRPRLLLGFAFFV 94

Qy 111 LICVAVIVVGFSSDIAAGLDYKHCSLYHGRPHAAVLYVGLFWLLDFSNNTVQGPARA 170
Db 95 ILAILCFTTPDFSDMGKLI-----YAYITYV---GLSLTYTTINVPYGA 135

Qy 171 MMADLC-DHHGPSAANSIFCSWALGNIL-----GYSSGSPNN-----WHKWFPELK 216
Db 136 LTSAMTRNNQEVSVTSVRMLFANLGLLVAVFAFFVPLLAAYLSDTSGNESLQMTMGLG 195

Qy 217 TSACCACANLKGAFLVAVVFLVLCV-----TVTLIFAKE-----VPRANENLPTT 263
Db 196 MIGGC-----LLIFCFKSTKRVTLQKSEEKIKFTDIFEQFRVNR----- 235
```

CC 264 KAGGEVETPTGLAVLKGFDLPFGMPVSLVLTAVTWSL-----WP----- 305
CC 236 -----PLVLSIFIIIFGVNSISVGVYVYTLNLERDLVKWYGLIGSLPA 283
CC 306 -----PFLYDMDGRIYHGPDKGSKNAQISAFNEGVGVGAGLLNSVILGFSFLIEP 361
CC 284 LVLPFIPRLHQLGK-----KLLNVALLN--IIGLLALLFVP 321
CC 362 MCRKGVPRVWVTSNFMVCMVMAAATLIS-----FWSL---RDYHGVODAITANASIKAV 414
CC 322 -----PSNVYL---LLVCLRLTAAGSLTAGGYMVALIPETIEYGYRTGKRMGGLIYAI 372
CC 415 CLVLRAF---LGVPLAIIYSVPFANTQAATRGGGQGLCTGVNLISIVIPQVIAL 468
CC 373 IGFFFKFGMALGVVPGVLDRFGVYVANAQT---PAALMGILITTTIPVFLVLV 425

RESULT 24
TAT2_YEAST
AC P38967; STANDARD: PRT; 592 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Tryptophan permease (Tryptophan amino acid transporter).
GN TAT2 OR TAP2 OR SCM2 OR LTC3 OR SAB2 OR YOL020W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=JK9-3D;
RX MEDLINE=95021187; PubMed=7523855;
RA Schmidt A., Hall M.N., Koller A.;
RT "Two FK506 resistance-conferring genes in Saccharomyces cerevisiae,
RT TAT1 and TAT2, encode amino acid permeases mediating tyrosine and
RT tryptophan uptake.";
RL Mol. Cell. Biol. 14:6597-6606(1994).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=94335876; PubMed=8058037;
RA Chen X.H., Xiao Z., Fitzgerald-Hayes M.;
RT "SCM2, a tryptophan permease in Saccharomyces cerevisiae, is
RT important for cell growth.";
RL Mol. Gen. Genet. 244:260-268(1994).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=YN 140;
RA Kawamura D., Yamashita I., Nimi O., Toh-E A.;
RT "Cloning and nucleotide sequence of a gene conferring ability to grow
RT at a low temperature on Saccharomyces cerevisiae tryptophan
RT auxotroph.";
RL J. Ferment. Bioeng. 77:1-9(1994).
RN
RP SEQUENCE FROM N.A.
RA Shin Y.H., Goo D.M., So I.S., Rhode P.R., Campbell J.L., Kim J.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE OF 1-481 FROM N.A.
RA Habbig B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE OF 156-592 FROM N.A.
RA Hughes B., Pohl T.M.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: REQUIRED FOR HIGH-AFFINITY TRYPTOPHAN TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL: X79150; CAA55777.1; -;
CC EMBL: L33461; AAA60324.1; -;
CC EMBL: D16304; BAA03811.1; -;
CC EMBL: U66834; AAB07526.1; -;
CC EMBL: Z74761; CAA99019.1; -;
CC EMBL: Z74762; CAA99020.1; -;
CC PIR: S48084; S48084.
CC PIR: S47926; S47926.
CC SGD: S0005380; TAT2.
CC InterPro: IPR002293; AA/rel_pmease1.
CC InterPro: IPR004840; AAC_permease.
CC InterPro: IPR004841; Permease.
CC InterPro: IPR004762; Yeast_AA_perm.
CC Pfam: PF00324; aa_permeases; 1.
CC TIGRfams: TIGR00913; 2A0310; 1.
CC PROSITE: PS00218; AMINO_ACID_PERMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 227 247 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
FT TRANSMEM 307 327 POTENTIAL.
FT TRANSMEM 359 379 POTENTIAL.
FT TRANSMEM 405 425 POTENTIAL.
FT TRANSMEM 429 449 POTENTIAL.
FT TRANSMEM 473 493 POTENTIAL.
FT TRANSMEM 515 535 POTENTIAL.
FT CONFLICT 187 187 Q -> T (IN REF. 3).
SQ SEQUENCE 592 AA; 65404 MW; A233BAD16264B319 CRC64;

Query Match 4.2%; Score 114.5; DB 1; Length 592;
Best Local Similarity 20.7%; Pred. No. 0.46;
Matches 104; Conservative 60; Mismatches 138; Indels 201; Gaps 28;

QY 25 HVAITSGLRLILAGNVAG-----GVQYGNALQSLTLTPVQTLGLSHALTSEFW 73
DB 86 HLIMTAIGSGTGTGLFVSGKRAIAGGLGVVIGWAIAGS-----QIGTIHGL----- 134
QY 74 LCGPIAGLVQPLVGLYSDRCTS-----RWGRRRPFILTCMLICVAVI 117
DB 135 --GEIT--VRPVGAFANYGTRFLDPSISFWSTIYVLQW-----FFVLPLEIAAAMT 185
QY 118 VVGFSSTGALGDTKEHCSLYHGPWHAIVY-----VLGFLLDFSNVTVOGP 167
DB 186 VOYMNSSIDPVI-----WVAIFYAVIVSINLFGVGRGGEAEFAFTIKAI 230
QY 168 ARAMADLC-----DH-----HGPSAANSIFCSMMALGNILGYS----- 201
DB 231 TVCGFIILCVVLICGGPDHPEIGAKYWHDPGCLANGEPGVLSVLVASYSLGGTEMCL 290
QY 202 --SGSTNNHKKWPFPLKTSACCEACANLKGAFVAVVFLVLCITVTLIFAKEVPYRANENL 260
DB 291 ASGETD--PKGLP-----SAIKQVFWRLFFFLISLTLVGFL--VPY-TNQL 333
QY 261 PTTKAGGEVETPTGLAVLKGFDLPFGMPVSLVLTAVTWSL----- 298
DB 334 ---LGGSSVDNSPFVIAIKLHHIKALPSIVNAVILSVLSVSGNSCIFASSRTLCSMAHQG 390
QY 299 -ITWLSWEPFLIYDFDWMGREIYHCDPKGSNAQISAFNEGVGVGAGLLNSVILGFSF 357
DB 391 LIPW--WFGYI---DRAGRPL-----VGMANS-LFGLLAF 420
QY 358 LIETPCMRKGVPRVWVTSNFMVCMVMAAATLISFW-SLRDHYGVQDAITANAS----- 410
DB 421 LV-----KSGS--MSEVENLWMAIAGLATCIV--WLSINLSHIFRLAMKAGKSLDELE 471

```
Qy 411 -IKAV-----CLVLF 420
: ||
Db 472 FVSAVGINGSVAVSALINCLILIA 494

RESULT 25
YJIO_ECOLI
ID YJIO_ECOLI STANDARD: PRT: 410 AA.
AC P39386;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transport protein yjio.
GN YJIO OR B4337.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=562;
RN [1]
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -1- SIMILARITY: STRONG, TO E.COLI MDFA.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U14003; AAA97233.1;
CC EMBL; AE000504; AAC77293.1;
CC EcoGene; EG12576; yjio.
CC Pfam; PF00083; sugar.tr.1.
CC Hypothetical protein; Transport; Transmembrane; Inner membrane;
CC Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 79 99 POTENTIAL.
FT TRANSMEM 104 124 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 217 237 POTENTIAL.
FT TRANSMEM 252 272 POTENTIAL.
FT TRANSMEM 283 303 POTENTIAL.
FT TRANSMEM 308 328 POTENTIAL.
FT TRANSMEM 349 369 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
SQ SEQUENCE 410 AA; 44687 MW; 17C909971AD788B7 CRC64;
```

```
Query Match 4.2%; Score 113.5; DB 1; Length 410;
Best Local Similarity 19.9%; Pred. No. 0.38;
Matches 100; Conservative 82; Mismatches 160; Indels 161; Gaps 29;

Qy 44 VOYGNALQSLTTPVOTGLSHALTSP-----MWLCGPYIAGLVVQPLVGLYS 91
: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19 ILYDFAAVLS--TDLIQP-GIINVVRDNADVSLAPAAVSLYLAG---GMALQWLLGPLS 72

Qy 92 DRCSTRWGRRRPFLTCGMLICVAVIVVGFSSD-----IGALGDTKEHCSLYHGPRWHA 146
: | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 73 DRI-----GRRPVLITGALIFTLACAATMTTTSMTQFLIARAIOGT-SICFI-----A 119
```

Search completed: April 17, 2003, 11:30:39
Job time : 35 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model.

Run on: April 17, 2003, 11:30:00 ; Search time 91 Seconds
(without alignments)
1175.148 Million cell updates/sec

Title: US-09-679-687A-2

Perfect score: 2731

Sequence: 1 MARGDGGQLAELSAAGRGAA.....VFLPKISKRQRAVSAGGH 519

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_21.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rviris.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2552	93.4	521	10 Q9SXM0	Q9SXM0 zea mays (m
2	2325	85.1	538	10 Q9LKH3	Q9LKH3 oryza sativ
3	2317.5	84.9	537	10 Q49838	Q49838 oryza sativ
4	2296.5	84.1	523	10 Q9M422	Q9M422 hordeum vul
5	2295.5	84.1	522	10 Q8S4W1	Q8S4W1 triticum ae
6	2287	83.7	523	10 Q8RUL3	Q8RUL3 triticum ae
7	2282.5	83.6	522	10 Q8S4W0	Q8S4W0 triticum ae
8	1862	68.2	506	10 Q948L0	Q948L0 oryza sativ
9	1862	68.2	506	10 Q944W2	Q944W2 oryza sativ
10	1551	56.8	612	10 Q9SQK5	Q9SQK5 vitis vinif
11	1549.5	56.7	604	10 Q9FVL6	Q9FVL6 lycopersico
12	1523	55.8	594	10 Q8O605	Q8O605 arabidopsis
13	1312.5	48.1	297	10 Q8S7X2	Q8S7X2 oryza sativ
14	1144	41.9	533	10 Q41152	Q41152 ricinus com
15	1126	41.2	501	10 Q65929	Q65929 daucus caro
16	1122.5	41.1	512	10 Q9S7Z5	Q9S7Z5 apium grave

Q9FE59 arabidopsis
Q8VYX4 brassica ol
Q9ZTB9 apium grave
Q9M3R4 arabidopsis
Q4Q938 plantago ma
Q9XHL6 pisum sativ
Q9SP63 vitis vinif
Q9SLN7 daucus caro
Q65803 daucus caro
Q9FNr6 daucus caro
Q4Q077 vicia faba
Q9FG00 arabidopsis
Q9FV92 solanum tub
Q9SP14 alonsoa mer
Q9ZVK6 arabidopsis
Q9Z232 arabidopsis
Q9FVJ6 lycopersico
Q9SQK6 vitis vinif
Q9C6H8 arabidopsis
Q9Z231 arabidopsis
Q9F1X9 arabidopsis
Q8O550 arabidopsis
Q9M423 hordeum vul
Q8RWQ6 arabidopsis
Q8VYX3 brassica ol
Q43653 solanum tub
Q9M535 euphorbia e
Q4Q937 plantago ma
Q9C8X2 arabidopsis
Q9SQK4 vitis vinif
Q9FNY9 arabidopsis
Q4O583 nicotiana t
Q94G12 oryza sativ
Q9SP15 asarina bar
Q39438 beta vulgar
Q9SM24 beta vulgar
Q04516 arabidopsis
Q9XFM1 nicotiana t
Q4O167 lycopersico
Q8W534 vitis vinif
Q65883 ricinus com
Q9AVR7 ricinus com
Q8W401 nicotiana t
Q94Jm9 oryza sativ
Q94Jm8 oryza sativ
Q9SVC5 betula verr
Q9SVV1 drosophila
Q90Z74 oryzias lat
Q14091 schizosacch
Q9HEX4 pneumocysti
Q9A612 caulobacter
Q961T2 homo sapien
Q95K15 macaca fasc
Q9SEAL apium grave
Q9SKC5 macaca fasc
Q9JSW6 neisseria m
Q9K111 neisseria m
Q9RVB3 deinococcus
Q8R7F4 thermoaer
Q9U060 schizosacch
Q9C716 arabidopsis
Q9V209 pyrococcus
Q9SXU7 cicer ariet
Q8ZTJ0 pyrococcus
Q9C1N0 penicillium
Q9ULN3 homo sapien
Q9V210 pyrococcus
Q9A5Y2 caulobacter
P70187 homo musculu
Q96NY0 homo sapien
Q96MC6 homo sapien
Q9DBS0 mus musculu
Q9Y2W3 homo sapien

90 140.5 5.1 448 16 Q8RGK4
 91 139 5.1 508 2 Q93KC2
 92 138 5.1 465 10 Q9MAN5
 93 137.5 5.0 447 2 Q03095
 94 137.5 5.0 473 17 Q97213
 95 136.5 5.0 394 16 Q92479
 96 136.5 5.0 789 16 Q8UGX6
 97 136 5.0 464 11 Q9DC37
 98 136 5.0 558 3 Q9UVM5
 99 133.5 4.9 399 16 Q8XSH5
 100 133.5 4.9 548 3 Q9P3K6

ALIGNMENTS

RESULT 1
 Q9SXM0
 ID Q9SXM0 PRELIMINARY; PRT; 521 AA.
 AC Q9SXM0;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Sucrose transporter.
 GN ZMSUT1.
 OS zea mays (Maize).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
 OC Panicoideae; Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Aoki N., Hirose T., Takahashi S., Ono K., Ishimaru K., Ohsugi R.;
 RT Molecular cloning and expression analysis of a gene for a sucrose
 RT transporter in maize (Zea mays L.)";
 RL Plant Cell Physiol. 0:0-0(1999).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL; AB008464; BAA83501.1;
 DR InterPro; IPR000847; HTH_LysR.
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 DR TIGRFAMs; TIGR01301; GPH_sucrose; 1.
 DR PROSITE; PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
 KW Transmembrane.
 SQ SEQUENCE 521 AA; 55174 MW; 50E38A5983B4F881 CRC64;

Query Match 93.4%; Score 2552; DB 10; Length 521;
 Best Local Similarity 92.4%; Pred. No. 2.7e-189;
 Matches 484; Conservative 16; Mismatches 16; Indels 8; Gaps 3;
 QY 1 MARGDGGQLAELSAGVRG-----AAAVVDHVAPISLGRLLAGMVGAGVQYGMALQSL 55
 DB 1 MARGDGE--LELSVGRGTGGGAAAAADHAPISLGRLLAGMVGAGVQYGMALQSL 58
 QY 56 TPVYQTLGLSHALTSEFMILCGPIAGLVQPLVGLYSDRCTSRWRRRPFILTGMLICVA 115
 DB TPVYQTLGLSHALTSEFMILCGPIAGLVQPLVGLYSDRCTSRWRRRPFILTGMLICVA 118
 QY 116 VIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSNVVGPARAMADL 175
 DB VIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSNVVGPARAMADL 178
 QY 176 CDHGHPSAANSIFCSWMALGNILGYSSGSTNNHKKWFFFLKTSACCEACANLKGALVAV 235
 DB CDHGHPSAANSIFCSWMALGNILGYSSGSTNNHKKWFFFLKTSACCEACANLKGALVAV 238
 QY 236 VFLVLCVLTFLIFAKEVYPYRANENLPTTKAGEVETETGPLAVLKGFDPKLPMPSPVLL 295
 DB VFLVLCVLTFLIFAKEVYPYRANENLPTTKAGEVETETGPLAVLKGFDPKLPMPSPVLL 297
 QY 296 VTATLWSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVYRGVAFGLLLNSVILGFS 355
 DB VTATLWSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVYRGVAFGLLLNSVILGFS 357

QY 356 SFLEPMDCKVGPVWVWTSNFMVCMVMAATALISFWSLRDHYGYVODAITANASIKAVC 415
 DB SFLEPMDCKVGPVWVWTSNFMVCMVMAATALISFWSLRDHYGYVODAITANASIKAVC 417
 QY 416 LVLFAPLGLVPLAILYSVPFAVTAQLAATRGGGGLCTGVLNLSIVIPQVITIALGAGPMDA 475
 DB LVLFAPLGLVPLAILYSVPFAVTAQLAATRGGGGLCTGVLNLSIVIPQVITIALGAGPMDA 477
 QY 476 LFGKGNTPAGVASAFALVGGVGVFLLPKISKRQFRAYSAGGH 519
 DB LFGKGNTPAGVASAFALVGGVGVFLLPKISKRQFRAYSAGGH 521

RESULT 2

Q9LKH3
 ID Q9LKH3 PRELIMINARY; PRT; 538 AA.
 AC Q9LKH3;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Sucrose transporter.
 GN SUT1.
 OS Oryza sativa (indica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartioideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39946;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. IR36;
 RA Whitfield P.R.; Furbank R.T.;
 RT "Sucrose transporter gene from rice (Oryza sativa indica cv. IR36).";
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF280050; AAF90181.1;
 DR InterPro; IPR003662; sub_transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 DR TIGRFAMs; TIGR01301; GPH_sucrose; 1.
 SQ SEQUENCE 538 AA; 56157 MW; 8DC0E7712A2167A1 CRC64;

Query Match 85.1%; Score 2325; DB 10; Length 538;
 Best Local Similarity 81.2%; Pred. No. 1e-171;
 Matches 436; Conservative 39; Mismatches 42; Indels 20; Gaps 3;

QY 1 MARGD-----GGQALAELSAGVRG-----AAAVVDHVAPISLGRLLAGMVA 41
 DB 1 MARGSGAGGGGGGGGLELSVGVGGGGGARGGGGGAATAAPISLGRLLSGMVA 60
 QY 42 GGVOYGMALQSLTTPVQTLGLSHALTSEFMILCGPIAGLVQPLVGLYSDRCTSRWRR 101
 DB GGVOYGMALQSLTTPVQTLGLSHALTSEFMILCGPIAGLVQPLVGLYSDRCTSRWRR 120
 QY 102 RPELTGCMILICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSN 161
 DB RPELTGCMILICVAVIVVGFSSDIGAALGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSN 180
 QY 121 RPYLTGCVLICLVAVVIGFSADIGYAMGDTKEDCSYHGSRWHAIVVVLGFWLLDFSN 221
 DB NTVOGPARAMADLCDHGHPSAANSIFCSWMALGNILGYSSGSTNNHKKWFFFLKTSACC 240
 QY 222 EACANLKGALVAVVFLVLCVLTFLIFAKEVYPYRANENLPTTKAGEVETETGPLAVLK 281
 DB EACANLKGALVAVVFLVLCVLTFLIFAKEVYPYRANENLPTTKAGEVETETGPLAVLK 299
 QY 282 GFKDLPMPSPVLLVTATLWSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVYRGV 341
 DB GFKDLPMPSPVLLVTATLWSWFFPILYDWMGREIYHGDPKGSNAQISAFNEGVYRGV 359
 QY 342 AFGLLLSNVILGFSFLLIEPMCKRKVGPVWVWTSNFMVCMVMAATALISFWSLRDHYGY 401
 DB AFGLLLSNVILGFSFLLIEPMCKRKVGPVWVWTSNFMVCMVMAATALISFWSLRDHYGY 419
 QY 402 QDAITANASIKAVCLVLAFLVGLVPLAILYSVPFAVTAQLAATRGGGGLCTGVLNLSIVI 461

Db 420 QKAITADKSKAVCLVFAELGVPLAVLYSVFPAVTAQLAATRGGGGGLCTGVLSNISIVI 479
 Qy 462 POVTIAGAGPMDALFCKGNIPAFGVASAFALGVGVGVVGVVGLLPKISKROFRAVSAGG 518
 Db 480 PQVIALGAGPMDLFGKGNIPAFGLASGAFALIGGVAGIFLLPKISKROFRAVSAGG 536

RESULT 3
 ID 049838 PRELIMINARY; PRT; 537 AA.
 AC 049838;
 DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Sucrose transporter.
 GN OSSU1.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE; TISSUE=GREEN LEAF;
 RX MEDLINE=98182940; PubMed=9522469;
 RA Hirose T., Imaizumi N., Scofield G.N., Furbank R.T., Ohsugi R.;
 RT "cDNA cloning and tissue specific expression of a gene for sucrose
 transporter from rice (Oryza sativa L.).";
 RL Plant Cell Physiol. 38:1389-1398(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL; D87819; BAA24071.1;
 DR InterPro; IPR003662; sub.transporter.
 DR Pfam; PF00083; sugar.tr.1.
 DR TIGRfams; TIGR01301; GPH_sucrose; 1.
 KW Transmembrane.
 SQ SEQUENCE 537 AA; 56130 MW; 4C1841581B810F84 CRC64;

Query Match 84.9%; Score 2317.5; DB 10; Length 537;
 Best Local Similarity 81.2%; Pred. No. 3.9e-171;
 Matches 435; Conservative 39; Mismatches 43; Indels 19; Gaps 3;

Qy 1 MARGD-----GGQLAELISAGVRG-----AAVVDHVAPISLGRLLIAGMVAG 42
 Db 1 MARGSGAGGGGGGGGGGGLSVGVGGGARGGGGGAATAAATPISLGRLLISGMVAG 60
 Qy 43 GVQYGNALQISLLTPYVQTGLSHALTSPFWLCPGPIAGLVVQPLVGLYSDRCTSRWGRRR 102
 Db 61 GVQYGNALQISLLTPYVQTGLSHALTSPFWLCPGPIAGLVVQPLVGLYSDRCTSRWGRRR 120
 Qy 103 PFILTCMLICVAVIVVGFSSDGAALGDTKEHCSLYRGRPRWAAIYVVLGFLLDFSN 162
 Db 121 PYILTGVCLICLAVVIGFADIGYANGDTKEDCSVTHGSRWAAIYVVLGFLLDFSN 180
 Qy 163 TVQGPARAMADCDHRGSAANSIFCSWMALGNILYSSGSTNNHKKWFPFLKTSACCE 222
 Db 181 TVQGPARAMADLSGRGPGTANSIFCSWMAMGNILYSSGSTNNHKKWFPFLKTRACCE 240
 Qy 223 ACANLKCAFLVAVVFLVCLTVTLIFAKEVPYRANENLPTTKAGGEVETPTGPLAVLKG 282
 Db 241 ACANLKCAFLVAVVFLVCLTVTLIFAKEVPYRANENLPTTKAGGEVETPTGPLAVLKG 299
 Qy 283 FKDLPPCMPSVLLVTAITLWSRPFILYDMDMGREIYHGDPKGSNAQISAFNEGVRVGA 342
 Db 300 FRNLPTCMPSVLIYVTLWSRPFILYDMDMGREIYHGDPKGTDQIEAFNGVRAGA 359
 Qy 343 FGLLSNVILGFSSFLIEPKRCVGRVWVTSNFWVCVMAAATALISFWSLRDHYGVQ 402
 Db 360 FGLLSNVILGFSSFLIEPKRCVGRVWVTSNFWVCVMAAATALISFWSLRDHYGVQ 419
 Qy 403 DAITANASIKAVCLVFAELGVPLAVLYSVFPAVTAQLAATRGGGGGLCTGVLSNISIVI 462
 Db 420 KAITADKSKAVCLVFAELGVPLAVLYSVFPAVTAQLAATRGGGGGLCTGVLSNISIVI 479

Qy 463 QVTIAGAGPMDALFCKGNIPAFGVASAFALGVGVGVVGVVGLLPKISKROFRAVSAGG 518
 Db 480 QVTIAGAGPMDLFGKGNIPAFGLASGAFALIGGVAGIFLLPKISKROFRAVSAGG 535

RESULT 4
 ID 09M422 PRELIMINARY; PRT; 523 AA.
 AC 09M422;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Sucrose transporter 1.
 GN SUT1.
 OS Hordeum vulgare (Barley).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticeae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. BARKE; TISSUE=DEVELOPING CARYOPSES;
 RA Weschke W., Panitz R., Sauer N., Wang Q., Neubohn B., Weber H.,
 RA Wobus U.;
 RT "Sucrose transport into Barley Seeds: Molecular Characterisation of
 two transporters and Implications for Seed Development and Starch
 Accumulation.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ272309; CAB75882.1;
 DR InterPro; IPR003662; sub.transporter.
 DR Pfam; PF00083; sugar.tr.1.
 DR TIGRfams; TIGR01301; GPH_sucrose; 1.
 SQ SEQUENCE 523 AA; 55251 MW; B53961C56365E9C0 CRC64;

Query Match 84.1%; Score 2296; DB 10; Length 523;
 Best Local Similarity 81.9%; Pred. No. 1.7e-169;
 Matches 429; Conservative 40; Mismatches 49; Indels 6; Gaps 2;

Qy 1 MARGDGGQLAELISAGVRG-----AAVVDHVAPISLGRLLIAGMVAGVQYGNALQISLL 55
 Db 1 MARGGGNGEVELSGVGGGGGAAPAAVQISLGRLLIAGMVAGVQYGNALQISLL 60
 Qy 56 TPVYQTGLSHALTSPFWLCPGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCMICVA 115
 Db 61 TPVYQTGLSHALTSPFWLCPGPIAGLVVQPLVGLYSDRCTSRWGRRRPFIITGCVLICLA 120
 Qy 116 VIVVGFSSDGAALGDTKEHCSLYRGRPRWAAIYVVLGFLLDFSNNTVOGPARAMADL 175
 Db 121 VIVVGFSSDGAALGDTKEHCSLYRGRPRWAAIYVVLGFLLDFSNNTVOGPARAMADL 180
 Qy 176 CDHGPSAANSIFCSWMALGNILYSSGSTNNHKKWFPFLKTSACCEACANLKCAFLVAV 235
 Db 181 SAQHGSAANSIFCSWMALGNILYSSGSTNNHKKWFPFLTRACCEACANLKCAFLVAV 240
 Qy 236 VFLVCLTVTLIFAKEVPYRANENLPTTKAGGEVETPTGPLAVLKGDKLPQMPSPVLL 295
 Db 241 LFLSLALVTLIFAKEVPYRANENLPTTKANGQVEVEPTGPLAVLKGDKLPQMPSPVLL 299
 Qy 296 VTATLWSRPFILYDMDMGREIYHGDPKGSNAQISAFNEGVRVAGFGLLSNVILGF 355
 Db 300 VTGLTWSRPFILYDMDMGREIYHGDPKGTPAANAFQEGVRAGFGLLSNVILGF 359
 Qy 356 SFLEPKRCVGRVWVTSNFWVCVMAAATALISFWSLRDHYGVQDATTANASIKAVC 415
 Db 360 SFLEPKRCVGRVWVTSNFWVCVMAAATALISFWSLRDHYGVQDATTANASIKAVC 419
 Qy 416 LVLFALGVPLAVLYSVFPAVTAQLAATRGGGGGLCTGVLSNISIVIPOVIALGAGPMDA 475
 Db 420 LALFALGVPLAVLYSVFPAVTAQLAATRGGGGGLCTGVLSNISIVIPOVIALGAGPMDA 479
 Qy 476 LFGKGNIPAFGVASAFALGVGVVGLLPKISKROFRAVSAGG 519

Db 480 LFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 523

RESULT 5

Q8S4W1 ID Q8S4W1 PRELIMINARY; PRT; 522 AA.

AC Q8S4W1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sucrose transporter SUT1A.

OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]

SEQUENCE FROM N.A.

RA Aoki N., Whitfield P.R., Hoeren F., Scofield G.N., Newell K.,
RA Patrick J., Offler C., Clarke B., Rahman S., Furbank R.T.;
RT "Three sucrose transporter genes are expressed in the developing grain
of wheat."
RL Plant Mol. Biol. 0:0-0(2002).
DR EMBL: AF408842; AAM13408.1; -;
SQ SEQUENCE 522 AA; 55072 MW; C980802E5593A6A CRC64;

Query Match 84.1%; Score 2295.5; DB 10; Length 522;

Best Local Similarity 81.2%; Pred. No. 1.9e-169;

Matches 428; Conservative 41; Mismatches 45; Indels 13; Gaps 3;

Qy 1 MARGDGGQLAELSAGVRGAAA-----VVDHVAPISLGRLLILAGMVAGGVQYGMALQL 52

Db 1 MARGGGNGEVELSVGCGGGAGGADAPVD---ISLGRLLILAGMVAGGVQYGMALQL 56

Qy 53 SLITPYVOTLGLSHALTSPFWLWCGPIAGLVVQPLVGLYSRDRCTSRWGRRRRPFILTGCMLI 112

Db 57 SLITPYVOTLGLSHALTSPFWLWCGPIAGLVVQPLVGLYSRDRCTSRWGRRRRPFILTGCLI 116

Qy 113 CVAVIVVGFSSDIGAALGDKHECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMM 172

Db 117 CIAVVVGFSSDIGAALGDKHECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMM 176

Qy 173 ADLCHGHPGSAANSIFCSWMAALGNILGYSRDRCTSRWGRRRRPFILTGCMLI 232

Db 177 ADLSAQHGHPGSAANSIFCSWMAALGNILGYSRDRCTSRWGRRRRPFILTGCMLI 236

Qy 233 VAVVFLVCLTVTLIFAKEVPRANENLPTTKAGGEVETPTGFLAVLKGFKDLPPGMP 292

Db 237 VAVVFLVCLTVTLIFAKEVPRANENLPTTKAGGEVETPTGFLAVLKGFKDLPPGMP 295

Qy 293 VLLVTAITWLSWFPFILDYDMGRIYHGDPKGSNAQISAFNEGVRVGFGLLINSVIL 352

Db 296 VLLVTAITWLSWFPFILDYDMGRIYHGDPKGSNAQISAFNEGVRVGFGLLINSVIL 355

Qy 353 GFSSFLIPEMCKKGVPRVWVTSNFMVCMATAALISFWSLRDYHGVDQAITANASIK 412

Db 356 GFSSFLIPEMCKKGVPRVWVTSNFMVCMATAALISFWSLRDYHGVDQAITANASIK 415

Qy 413 AVCLVLEAFGLVPLAILYVFPFAVTAQAATRGGGGLCTGVNLISIVIPQVITAGAGP 472

Db 416 IVSLALFAFLGIPLAILYVFPFAVTAQAATRGGGGLCTGVNLISIVIPQVITAGAGP 475

Qy 473 WDALFKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 519

Db 476 WDELFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 522

RESULT 6

Q8RUL3

ID Q8RUL3 PRELIMINARY; PRT; 523 AA.

AC Q8RUL3;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DE Sucrose transporter SUT1B.

OS Triticum aestivum (wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;

OC Triticeae; Triticum.

OX NCBI_TaxID=4565;

RN [1]

SEQUENCE FROM N.A.

RA Aoki N., Whitfield P.R., Hoeren F., Scofield G.N., Newell K.,

RA Patrick J., Offler C., Clarke B., Rahman S., Furbank R.T.;

RT "Three sucrose transporter genes are expressed in the developing grain

of wheat."
RL Plant Mol. Biol. 0:0-0(2002).
DR EMBL: AF408844; AAM13410.1; -;
SQ SEQUENCE 523 AA; 55232 MW; AD42EBDB3B7F734F CRC64;

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Sucrose transporter SUT1D.

OS Triticum aestivum (wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;

OC Triticeae; Triticum.

OX NCBI_TaxID=4565;

RN [1]

SEQUENCE FROM N.A.

RA Aoki N., Whitfield P.R., Hoeren F., Scofield G.N., Newell K.,

RA Patrick J., Offler C., Clarke B., Rahman S., Furbank R.T.;

RT "Three sucrose transporter genes are expressed in the developing grain

of wheat."
RL Plant Mol. Biol. 0:0-0(2002).
DR EMBL: AF408845; AAL90455.1; -;
DR EMBL: AF408844; AAM13410.1; -;
SQ SEQUENCE 523 AA; 55232 MW; AD42EBDB3B7F734F CRC64;

Query Match 83.7%; Score 2287; DB 10; Length 523;

Best Local Similarity 81.1%; Pred. No. 8.6e-169;

Matches 425; Conservative 42; Mismatches 51; Indels 6; Gaps 2;

Qy 1 MARGDGGQLAELSAGVRG-----AAVVDHVAPISLGRLLILAGMVAGGVQYGMALQLSLL 55

Db 1 MARGGGNGEVELSVGCGGGAGGGEQAPAVDISLGRLLILAGMVAGGVQYGMALQLSLL 60

Qy 56 TPYVOTLGLSHALTSPFWLWCGPIAGLVVQPLVGLYSRDRCTSRWGRRRRPFILTGCMLCVA 115

Db 61 TPYVOTLGLSHALTSPFWLWCGPIAGLVVQPLVGLYSRDRCTSRWGRRRRPFILTGCMLCIA 120

Qy 116 VIVVGFSSDIGAALGDKHECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 175

Db 121 VIVVGFSSDIGAALGDKHECSLYHGPRWHAIVVVLGFWLLDFSNNTVQGPARAMMADL 180

Qy 176 CDHGHGHPGSAANSIFCSWMAALGNILGYSRDRCTSRWGRRRRPFILTGCMLCVA 235

Db 181 SAQHGHPGSAANSIFCSWMAALGNILGYSRDRCTSRWGRRRRPFILTGCMLCVA 240

Qy 236 VFLVCLTVTLIFAKEVPRANENLPTTKAGGEVETPTGFLAVLKGFKDLPPGMP 295

Db 241 LFLAFLVTLVTLIFAKEVPRANENLPTTKAGGEVETPTGFLAVLKGFKDLPPGMP 299

Qy 296 VTATITWLSWFPFILDYDMGRIYHGDPKGSNAQISAFNEGVRVGFGLLINSVILGFS 355

Db 300 VTGLTWLSWFPFILDYDMGRIYHGDPKGSNAQISAFNEGVRVGFGLLINSVILGFS 359

Qy 356 SFLIEPMCKKGVPRVWVTSNFMVCMATAALISFWSLRDYHGVDQAITANASIKAVC 415

Db 360 SFLIEPMCKKGVPRVWVTSNFMVCMATAALISFWSLRDYHGVDQAITANASIKAVC 419

Qy 416 LVLEAFGLVPLAILYVFPFAVTAQAATRGGGGLCTGVNLISIVIPQVITAGAGPMDA 475

Db 420 LVLEAFGLVPLAILYVFPFAVTAQAATRGGGGLCTGVNLISIVIPQVITAGAGPMDA 479

Qy 476 LFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 519

Db 480 LFGKGNIPAFGMSVFALIGGVWGIPLLPKISRQFRVAVSGGH 523

RESULT 7

Q8S4W0 ID Q8S4W0 PRELIMINARY; PRT; 522 AA.

AC Q8S4W0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DE Sucrose transporter SUT1B.

OS Triticum aestivum (wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;

OC Triticeae; Triticum.

OX NCBI_TaxID=4565;

RN [1]

RP SEQUENCE FROM N.A.
RA Aoki N., Whitfield P.R., Hoeren F., Scofield G.N., Newell K.,
RA Patrick J., Offler C., Clarke B., Rahman S., Furbank R.T.,
RT "three sucrose transporter genes are expressed in the developing grain
of wheat.";
RL Plant Mol. Biol. 0:0-0(2002).
DR EMBL: AF408843; AAL13409.1; -;
SQ SEQUENCE 522 AA; 55168 MW; 20EC33438765D734 CRC64;

Query Match 83.6%; Score 2282.5; DB 10; Length 522;
Best Local Similarity 81.0%; Pred. No. 1.9e-188;
Matches 421; Conservative 41; Mismatches 46; Indels 13; Gaps 3;

Qy 1 MARGGGGGLAELASAGVGA---VVDHVAPISLGRLLILAGMVAGGVQVGMALQL 52
Db 1 MARGGGGGEVLSVGGGGAGAGADAPVD---ISLGRLLILAGMVAGGVQVGMALQL 56

Qy 53 SLLTPYVQTGLSHALTSFMWLCGPIAGLVVQPLVGLYSRDRCTSRWGRRRPFIITGCM 112
Db 57 SLLTPYVQTGLSHALTSFMWLCGPIAGLVVQPCVGLYSRDRCTSRWGRRRPFIITGY 116

Qy 113 CVAVTVGFSDDIGAALGDTREHCSLYHGPRWHAIVVVLGFWLLDFSNNTVOGPARAM 172
Db 117 CIAVVVGFSADIGAALGDSKECSLYHGPRWHAIVVVLGFWLLDFSNNTVOGPARAL 176

Qy 173 ADLCHDHGSPSAANSIFCSWALGNILGYSSGSTNNHKKWFFPKTSACCEACANLKGAFL 232
Db 177 ADLSAQHGSPSAANSIFCSWALGNILGYSSGSTNNHKKWFFPKTRACCEACANLKGAFL 236

Qy 233 VAVVFLVLCVLTIFAKEVYPRANENLPTTKAGEVEETPTGPLAVLKGFDPKPLPGMPS 292
Db 237 VAVVFLVLCVLTIFAKEVYPRANENLPTTKAGEVEETPTGPLAVLKGFDPKPLPGMPS 295

Qy 293 VLLVTAITWLSWFFELVDTDMGREIYHGDPKGSNAQISAFNEGVRVGFAGGLLNSVIL 352
Db 296 VLLVTAITWLSWFFELVDTDMGREIYHGDPKGSNAQISAFNEGVRVGFAGGLLNSVIL 355

Qy 353 GFSSFLIEPMCKRKVPRVYVNTSMFVMAAATLISFWSLDYHGVDVDAITANASIK 412
Db 356 GFSSFLIEPLCKRLKLPVWVSSNLFVLSMAAIIISWMAATDMGHYQIHAIKREIK 415

Qy 413 AVCLVLEAFGLVPLAIVSFVNTAQLAATRGGGGGLCTGVLNIVIPQVITAGL 472
Db 416 IVSLLAFGLVPLAIVSFVNTAQLAATRGGGGGLCTGVLNIVIPQVITAGL 475

Qy 473 WDLFGKGNIPAFGVASAFALVGGVGVFLPKISKROFRAVSAGGH 519
Db 476 WDLFGKGNIPAFGVASAFALVGGVGVFLPKISKROFRAVSAGGH 522

RESULT 8
Q948L0 ID Q948L0 PRELIMINARY; PRT; 506 AA.
AC Q948L0
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sucrose transporter.
GN OSSUT3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE; TISSUE=PANICLE;
RA Hirose T., Scofield G.N., Whitfield P.R., Aoki N., Furbank R.T.,
RA Terao T.;
RT "Isolation and characterisation of a cDNA for a novel sucrose
transporter, OSSUT3, from rice.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB071809; BAB68368.1; -;

DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRFAMs; TIGR01301; GPH_sucrose; 1.
SQ SEQUENCE 506 AA; 52776 MW; 0844DC10E1E63E75 CRC64;

Query Match 68.2%; Score 1862; DB 10; Length 506;
Best Local Similarity 70.3%; Pred. No. 6.1e-136;
Matches 362; Conservative 47; Mismatches 86; Indels 20; Gaps 6;

Qy 11 ELSAGVRCGAAAVDHDVAP--ISLGRLLILAGMVAGGVQVGMALQLSLLTPYVQTGLSHAL 68
Db 6 ELDDGGDGGK-----APQISLSGLFLACMVAGGVQVGMALQLSLLTPYVQTGLSHAL 60

Qy 69 TSFWMWLCGPIAGLVVQPLVGLYSRDRCTSRWGRRRPFIITGCMICVAVIVVGFSSDIAA 128
Db 61 TSFWMWLCGPIAGLVVQPCVGLYSRDRCTSLGRRRPFIITGCIICISIVIVGFSSDIGA 120

Qy 129 LGDTKEHCSLYHGPRWHAIVVVLGFWLLDFSNNTVOGPARAMADLCHDHGSPSAANSIF 188
Db 121 LGDTTDCVYRGPRYHAAAFAILGFWLLDFSNNTVOGPARALMADLSGRHGPSAANAIF 180

Qy 189 CSWALGNILGYSSGSTNNHKKWFFPKTSACCEACANLKGAFLVAVVFLVLCVLTIF 248
Db 181 CSWALGNILGYSSGSTNDHKKWFFPKTRACCEACANLKAAFLVAVVFLGLSTAVTWVF 240

Qy 249 AKVYPRANENLPTTKAGEVEETPTGPLAVLKGFDPKPLPGMPSVLLTAITWLSWFFPI 308
Db 241 AREVA-----LDPVAAAKRNEGEASGLLAVFKGNKLPVGMPSVLIITGLTWSWFFPI 294

Qy 309 LYDTDMGREIYHGDPKGSNAQISAFNEGVRVGFAGGLLNSVILGFSFLIEPMCKRKVGP 368
Db 295 LFDTDMGREIYHGRPDGPAEVTAQEGVRQAGFGLLNSIVLGISSFLIEPMCKRRLGA 354

Qy 369 RVVWVTSNFMVMAAATLISFWSLDYHGVDVDAITANA---SIKAVCLVLEAFGLV 425
Db 355 RAVWVSSAVVAVVMAAATLISFWSLDYHGVDVDAITANA---SIKAVCLVLEAFGLV 414

Qy 426 LAIVSVFVNTAQLAATRGGGGGLCTGVLNIVIPQVITAGLGPWDALFGKGNIPAF 485
Db 415 FAVLCSVFFVNTAQLAASRGGGGLCTGVLNIVIPQVITAGLGPWDALFGKGNIPAF 474

Qy 486 GVASAFALVGGVGVFLPKISKROFRAVS-AGGH 519
Db 475 AMASVFAAAGVAVVLLPKVS---VRSVSMAGGH 506

RESULT 9
Q944W2 ID Q944W2 PRELIMINARY; PRT; 506 AA.
AC Q944W2
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Sucrose transporter.
OS Oryza sativa (indica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39946;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. IR 36;
RA Whitfield P.R., Aoki N., Scofield G.N., Hirose T., Terao T.,
RA Furbank R.T.;
RT "Isolation and characterisation of a putative sucrose transporter gene
(OSSUT3) from rice.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF419298; AAL14982.1; -;
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRFAMs; TIGR01301; GPH_sucrose; 1.
SQ SEQUENCE 506 AA; 52774 MW; A4E73029D7022B64 CRC64;

```
Query Match 68.2%; Score 1862; DB 10; Length 506;
Best Local Similarity 69.9%; Pred. No. 6.1e-136;
Matches 360; Conservative 48; Mismatches 87; Indels 20; Gaps 6;

QY 11 ELSAGVRGAAVVDHAP--ISLGRLLIAGMVAGGVQVQGWALQSLTTPYVQTLGSHAL 68
DB 6 ELDDGGDGRK-----APPQISLGLFLACMVAGGVQVQGWALQSLTTPYIQTGLIPHAL 60

QY 69 TSFMLCGPIAGLVQPLVGLYSDRCTSRGRRRPFILTGCMLICVAVIVGVFSSDIGAA 128
DB 61 TSMWMLCGPIAGLIVQPCVGLYSDCTSLGRRRPFILTGCIICISIVIVGVFSSDIGVA 120

QY 129 LGDTKEHCSLYHGRWHAAIVVVLGFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANSIF 188
DB 121 LGDATEDCKVYRGRPRYHAAAFLIGFWLLDFSNNTVQGPARAMMADLCHDHGSPSAANIF 180

QY 189 CSWMLGNLIGYSSGNTNWNHKKWFPELKTSSACCEACANLKGAFVAVVFLVCLVTFLIF 248
DB 181 CSWMLGNLIGYSSGNTNWNHKKWFPELKTSSACCEACANLKAFLVAVVFLVGLSTAVTWVF 240

QY 249 AKFVPRANENLPTTKAGEVETETGPLAVLKGKDLPPGMPSVLLVTAITWLSWEPFI 308
DB 241 AREVA-----LDPVAAKRNKEGEASGPLAVFKGMKLPVGMPSVLLVITGLTWLSWEPFI 294

QY 309 LYDTDMGRIYHGPDKGNSAQISAFNEGVRYGAFGLLINSVILGFSSFLIEPDMCRKVP 368
DB 295 LFDTDMGRIYHGRPDGSPAETAFQEGVROGAFGLLINSVILGFSSFLIEPDMCRRLGA 354

QY 369 RVVWTSNPMVCVMAATALISFWSLRDYHGVQDAITANA----SIKAVCLVFLFAFLGVP 425
DB 355 RAVVWSSAVVCVMAAVSVLSAWSLGDGFGVQDAARAPAEQGVRSALALFVFLGIP 414

QY 426 LALLVSPFAVTAQLAATRGGGGLCTGVNLISIVIPQVITIALGAGPMDALFGKGNIPAF 485
DB 415 FAVLCSVPAVTAQLTASRGGGGLCTGVNLISIVIPQVITIALGAGPMDALFGKGNIPAF 474

QY 486 GVSALYGVGVGVVFLFKPKSKROFRAVS--AGGH 519
DB 475 AMASVFAAAAAAGVLLPKVS---VRSVSMAGH 506

RESULT 10
Q9SQK5 PRELIMINARY; PRT; 612 AA.
AC Q9SQK5;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE Putative sucrose transporter.
GN VWSUC12.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Vitis.
OX NCBI_TaxID=29760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHIRAZ; TISSUE=FRUIT;
RA Davies C., Wolf T., Robinson S.P.;
RT "Three putative sucrose transporters are differentially expressed in
RL grapevine tissues.";
RL Plant Cell 147:93-100(1999).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF021809; AAF08330.1;
DR TIGRfams; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 612 AA; 65711 MW; 8FC3FBD6AB439078 CRC64;

Query Match 56.8%; Score 1551; DB 10; Length 612;
Best Local Similarity 54.8%; Pred. No. 8.5e-112;
Matches 298; Conservative 78; Mismatches 110; Indels 58; Gaps 6;

QY 31 LGRLLIAGMVAGGVQVQGWALQSLTTPYVQTLGSHALTFSFWMLCGPIAGLVQPLVGLY 90
```

```
DB 63 LRTLSCMIAAGVQVQGWALQSLTTPYIQTGLGIEHAFSSFIWLCGPTIGLVQPCVGIW 122
QY 91 SDRCTSRGRRRPFILTGCMLICVAVIVGVFSSDIGALGDTKEHCSLYHGRWHAAIVY 150
DB 123 SDKCSSKTYGRRRPFILAGLSMISVAVTIIGFSADIGYLLGDTNMDCKRFKGTRTWAAIIF 182
QY 151 VLGFLLDFSNNTVQGPARAMMADLCHDHGSPSAANSIFCSWMLGNLIGYSSGNTNWNHKK 210
DB 183 VLGFWMLDLANTVQGPALLADLSGDPDRNSANAIFCSWMLGNLIGFSAGASGHWR 242
QY 211 WFPFLKTSACCEACANLKGAFVAVVFLVCLVTFLIFAKEVP----- 253
DB 243 WFPFLNKKACCEACGNLKAFLIIVVFLTCLTFLYFAEEVPLMAYQPHHLSDSAPLLD 302
QY 254 -----YRANEN-----LPTTKAGEVETETPGPLAY 279
DB 303 NPOQIGFONSCKLDMASVDNATGNPNPESSEYINKNAKHLTPVIOEQNESFSFGPGLV 362
QY 280 --LKGEKDLPPGMPSVLLVTAITWLSWFPFIYLDYDMGRIYHGPDKGNSAQISAFNEG 337
DB 363 NLLTSRLHLPAMHSVLLVMAISWLSWFPFLFDYDMGRIYHGPDKGNSAQISAFNEG 422
QY 338 VRVGAFLGLLSVILGFSSFLIEPDMCRKVPVWVTSNPMVCVMAATALISFWSLRDY 397
DB 423 VREGAFGLLSVILGFSSFLIEPDMCRKVPVWVTSNPMVCVMAATALISFWSLRDY 482
QY 398 --HGVQDAITANASIKAVCLVFLFAFLGVPFAVTAQLAATRGGGGLCTGV 455
DB 483 ITEG-IQHAIGENRAIKTASLVVFAFGFLPSITYVSPFISITAEITADTGGGGLAIGVL 541
QY 456 NISVIPQVITIALGAGPMDALFGKGNIPAFVAVVGVVGVVFLFKPKSKROFRAVS 515
DB 542 NLAIVIPQVITIALGAGPMDALFGKGNIPAFVAVVGVVGVVFLFKPKSKROFRAVS 599
QY 516 AGGH 519
DB 600 SGFH 603

...RESULT 11
Q9FVL6 PRELIMINARY; PRT; 604 AA.
AC Q9FVL6;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Sucrose transporter-like protein.
GN SUT2.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20359188; PubMed=10899981;
RA Barker L., Kuhn C., Weise A., Schulz A., Gebhardt C., Hirner B.,
RA Hellmann H., Schulze W., Ward J.M., Frommer W.B.;
RT "SUT2, a putative sucrose sensor in sieve elements.";
RL Plant Cell 12:1153-1164(2000).
DR EMBL; AF166498; AAG12987.1;
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr_1.
DR TIGRfams; TIGR01301; GPH_sucrose; 1.
SQ SEQUENCE 604 AA; 65082 MW; 3403349B1692B2AF CRC64;

Query Match 56.7%; Score 1549.5; DB 10; Length 604;
Best Local Similarity 54.8%; Pred. No. 1.1e-111;
Matches 298; Conservative 73; Mismatches 116; Indels 57; Gaps 5;

QY 30 SLGRLLIAGMVAGGVQVQGWALQSLTTPYVQTLGSHALTFSFWMLCGPIAGLVQPLVGL 89
DB 63 LRTLSCMIAAGVQVQGWALQSLTTPYIQTGLGIEHAFSSFIWLCGPTIGLVQPCVGIW 122
```

Db 62 SLTLILSLCTVAAGVQFGALQSLTLPYIQTGLIEHAFSSFIWLCGPITGLVVPQCVGI 121
 QY 90 YSDRCSRWGRRRPFIITGCMILICVAVIVGVFSSDICAALGDDTKHECHSLYHGPWHAAIV 149
 Db 122 WSDCKSHKYGRRRRPFIIGAVNLSIAVIIIGFSDADIGLLODTHKCHSTFGKTRRAIV 181
 QY 150 YVGFLLDSSNNTVQGPAMMADLCHDHGSPASANSIFCSWMAALNIGLYSGSTNNWH 209
 Db 182 FVGVFWMLDLNNTVQGPAMMADLCHDHGSPASANSIFCSWMAALNIGLYSGSTNNWH 241
 QY 210 KWPFLLKTSACCEACANLKGAFLVAVVFLVCLTVTLIFAKEVP----- 253
 Db 242 RWPFLLNRAACCEPCGNLKAFLVAVVFLVCLTVTLIFAKEVP----- 301
 QY 254 -----YR-----ANENLPTTKAGGEVETPTGPL 277
 Db 302 DSPONTGFDLSQSKRELQYNSVANNESEMCHVADNSPKNEEQRPDKDQDSDADSPGAV 361
 QY 278 AV--LKGFKDLPPGMPSVLLVTAITWLSWFFPFIYDTHMGREIYHGDPKGSNAQISAFN 335
 Db 362 LVNLLTSLRHLPPAMHSVLIVMALTWLSWFFPFIYDTHMGREIYHGDPKGEADVNAYN 421
 QY 336 EGVVGAAGLLNSVILGFSSEFLIEPCKRVGVVWVTSNFWVCVMAATALISFWSLR 395
 Db 422 QGVREGAFGLLLNSVILGFSSEFLIEPCKRWIGSLWAVSNFVFCMACTAIISVVIS 481
 QY 396 DYHGYVQDAITANASIKAVCLVLFAPLGVPLAILYSVPFAVTAQLAATRGGOGGLCTGV 455
 Db 482 ANTOGVQHVIGATSTOIALVFLSLGIPLVITYSVFPTAITELTADAGGOGGLAIGVL 541
 QY 456 NISTVIPQVIALGAGPMDALFGKGNIPAFGVASAFALVGVGVGVFLPKIKRQFRAYS 515
 Db 542 NLATVVPQMVVSLGAGPMDALFGKGNIPAFALASALAAGIIFAMRLPLNLS--S 598
 QY 516 AGGH 519
 Db 599 TGFH 602

RESULT 12
 ID 080605 PRELIMINARY; PRT; 594 AA.
 AC 080605;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Putative sucrose/H+ symporter (Sucrose transporter).
 GN T17M13.3 OR SUC3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Renning C.M., Fraser C.M., Somerville C.R., Venter J.C.,
 RT "Arabidopsis thaliana chromosome II BAC T17M13 genomic sequence.",
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Meyer S., Truernit E., Huemmer C., Besenbeck R., Stadler R., Sauer N.,
 RT "AtSUC3: an unusual sucrose transporter from Arabidopsis.",
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 DR EMBL; AC004138; AAC32907.1; -
 DR EMBL; AJ289165; CAB92307.1; -
 DR InterPro: IPR003662; sub.transmembrane.
 DR Pfam: PF00083; sugar.tr.1.
 DR TIGRfam; TIGR01301; GPH_sucrose; 1.
 KW Transmembrane.
 SQ SEQUENCE 594 AA; 63972 MW; DF075A3601299DA0 CRC64;

Query Match 55.8%; Score 1523; DB 10; Length 594;
 Best Local Similarity 53.6%; Pred. No. 1.2e-109;
 Matches 288; Conservative 82; Mismatches 117; Indels 50; Gaps 5;
 QY 30 SLGRLLIAGMVAGVQYGMALQSLTLPYIQTGLSHALTSFWMWLCGPITGLVVPQCVGI 89
 Db 59 SLTVLSCSTVAAGVQFGALQSLTLPYIQTGLISHAFSSFIWLCGPITGLVVPQCVGI 118
 QY 90 YSDRCSRWGRRRPFIITGCMILICVAVIVGVFSSDICAALGDDTKHECHSLYHGPWHAAIV 149
 Db 119 WSDCKTSKYGRRRRPFIIGAVNLSIAVIIIGFSDADIGLLODTHKCHSTFGKTRRAIV 178
 QY 150 YVGFLLDSSNNTVQGPAMMADLCHDHGSPASANSIFCSWMAALNIGLYSGSTNNWH 209
 Db 179 FVGFWMMLDLNNTVQGPAMMADLCHDHGSPASANSIFCSWMAALNIGLYSGSTNNWH 238
 QY 210 KWPFLLKTSACCEACANLKGAFLVAVVFLVCLTVTLIFAKEVPYRANE----- 258
 Db 239 EWPFLLTSRACCAACGNLKAFLVAVVFLVCLTVTLIFAKEIPTSFKTRIQDSAPLL 298
 QY 259 -----NLPTTKAG-----EVETPTGPLAY--LKG 282
 Db 299 DLQSKGLEHSLKLNGTANGIKYVERVDTDEQFGNSENEHQDETYVDGPGSVLNLTS 358
 QY 283 FKDLPPGMPSVLLVTAITWLSWFFPFIYDTHMGREIYHGDPKGSNAQISAFNEGVVGA 342
 Db 359 LRHLPPAMHSVLIVMALTWLSWFFPFIYDTHMGREIYHGDPTGDSLHMLYDQGVREGA 418
 QY 343 FGLLLNSVILGFSSEFLIEPCKRVGVVWVTSNFWVCVMAATALISFWSLRDYHGVVQ 402
 Db 419 LGLLLNSVILGFSSEFLIEPCKRMGARGVWALSNTVFACMAGTAVISLMSLSDKNGIE 478
 QY 403 DAITANASIKAVCLVLFAPLGVPLAILYSVPFAVTAQLAATRGGOGGLCTGVLSIVIP 462
 Db 479 YIMRGNETTAAVIVFALLGFLPILAITYSVPFVTAETADSGGGGLAIGVLNLAIVIP 538
 QY 463 QVIALGAGPMDALFGKGNIPAFGVASAFALVGVGVGVFLPKIKRQFRAYSAGGH 519
 Db 539 OMIVSLGAGPMDQLFGGNNLPFVLSVAAFAAGVIALQRLPTLSS-SFK--STGFH 592

RESULT 13
 ID 0857X2 PRELIMINARY; PRT; 297 AA.
 AC 0857X2;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE Sucrose transporter, 5'-partial (Fragment).
 GN OSJNBA0091P11.1.
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPONBARE;
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
 RA Gansberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
 RA Tsitrin T., Riggs F., Hsiao J., Zismann V., Blunt S., Pal G.,
 RA VanAken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
 RA Salzberg S.L., White O., Fraser C.M.;
 RT "Oryza sativa chromosome 3 BAC OSJNBA0091P11 genomic sequence.",
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC073556; AAL84308.1; -
 FT NON_TER 1
 SQ SEQUENCE 297 AA; 31370 MW; BD332D42C395F9D7 CRC64;

Query Match. 48.1%; Score 1312.5; DB 10; Length 297;
 Best Local Similarity 83.4%; Pred. No. 1e-93;
 Matches 247; Conservative 25; Mismatches 23; Indels 1; Gaps 1;

```
QY 223 ACANLKGAFVAVFLVLCITVTLIFAKEVPYRANENLPTTKAGGEVETETGTLAVLKG 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 ACANLKGAFVAVFLVLCITVTLIFAKEVPYRANENLPTTKAGGEVETETGTLAVLKG 59
QY 283 FKDLPPGMPVLLVTAITWLSWFFPILYDMDGRIYHGDGPKGNAQISAFNEGVRYGA 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 60 FRLNLTGMPVLLVTAITWLSWFFPILYDMDGRIYHGDGPKGNAQISAFNEGVRYGA 119
QY 343 FGLLNSVILGFSFLIEPDMCRKVGPRVWVTSNFMVCMVAMAATALISFWSLRDHYGYVO 402
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 120 FGLLNSVILGFSFLIEPDMCRKVGPRVWVTSNFMVCMVAMAATALISFWSLRDHYGYVO 179
QY 403 DAITANASIKAYCLVFLAFGLVPLAFLVPLAFLVPLAFLVPLAFLVPLAFLVPLAFLV 462
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 180 KAITADKSIKAYCLVFLAFGLVPLAFLVPLAFLVPLAFLVPLAFLVPLAFLVPLAFLV 239
QY 463 QVVIAGAGPMDALGKGNIPAFGVASAFALVGGVGVFLPLPKISKROFRAVSAGG 518
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 240 QVVIAGAGPMDALGKGNIPAFGVASAFALVGGVGVFLPLPKISKROFRAVSAGG 595

RESULT 14
Q41152 PRELIMINARY; PRT: 533 AA.
ID Q41152
AC Q41152
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Sucrose carrier.
OS SCRL.
GN Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CARMENCITA; TISSUE=COTYLEDON;
RA Weig A., Komor E.;
RT "An active sucrose carrier (Scrl) that is predominantly expressed in
the seedling of Ricinus communis L.";
RL J. Plant Physiol. 147:685-690(1996).
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; Z31561; CAA83436.1; -
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 533 AA; 56359 MW; 0643E25D525AAAD8 CRC64;

Query Match 41.9%; Score 1144; DB 10; Length 533;
Best Local Similarity 45.0%; Pred. No. 2.2e-80;
Matches 226; Conservative 99; Mismatches 151; Indels 26; Gaps 8;

QY 19 AAADVHVAPISLGRLLAGVAGVQYQGWALQLSLTPYVQTLGLSHALTSPFMLCGPI 78
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 25 AGAAENSSP--LRKVMVVASIAAGIQGFWALQLSLTPYVQTLGLSHALTSPFMLCGPI 82
QY 79 AGLVQPLVGLYSDRCTSRWRRRPFILITGCMILICVAVIVVGFSSDGAALGDTKEHCSL 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 83 SGMVLQPIVGVHSDRCTSRWRRRPFILITGCMILICVAVIVVGFSSDGAALGDTKEHCSL 137
QY 139 YHGPWRHAAIVVVLGFWLLDFSNNTVOGPARAMMADLCDDHGPS-----AANSIFCSNMA 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 138 DSKPKTRATATFVGVFWLLDVANMLQGPCRALLAD---SGTSOKKTRTANLSTFFMA 194
QY 194 LGNILGYSSGSTNNHKKPPFLTKTSACCEACANLKGAFVAVFLVLCITVTLIFAKEVP 253
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 195 VGNVLGYAAGAYTHLYKLPFPFTKTACDVCYCANLKSCEFISIVLLSLTVLALSIVKKEP 254
QY 254 Y---RANENLPTTKAGGEVETETGTL--AVLKGFKDLPPGMPVLLVTAITWLSWFFPI 308
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 255 WSPQADVNAEDDTASQASSAQPMPPFGEILGAFKNLKRPMWILLVTCNLIAWAFPEL 314
QY 309 LVDYDMGRIYHGDGPKGNAQISAFNEGVRYGAFGLLNSVILGFSFLIEPDMCRKV-G 367
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 315 LFDYDMGRIYHGDGPKGNAQISAFNEGVRYGAFGLLNSVILGFSFLIEPDMCRKV-G 374
QY 368 PRVWVTSNFMVCMVAMAATALI-----SFWSLRDHYGYVQDAITANASIKAYCLVFLAFL 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 375 VKRLWGVNFMVCMVAMAATALI-----SFWSLRDHYGYVQDAITANASIKAYCLVFLAFL 434
QY 423 GYPLAFLVSVFPAVTAQAALATRGQGLCTGVNLISIVIPQVITIALGAGPMDALGKGN 482
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 435 GVPQAITYSIFPALASIFNSGTSGAGQGLSLGVNLISIVIPQVITIALGAGPMDALGKGN 494
QY 483 PAFGVASAFALVGGVGVFLPL 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 495 PAFVGVAAALASGIFALTMLP 516

RESULT 15
O65929 PRELIMINARY; PRT: 501 AA.
ID O65929
AC O65929;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Sucrose/H+ symporter (Sucrose/proton symporter).
GN SUN1B OR SUN1A.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANTALISE; TISSUE=LEAVES;
RA Shakya R., Sturm A.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANTALISE;
RA Sturm A.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NANTALISE;
RA Shakya R.;
RL Thesis (2000), Department of Department of Botany,
University of Basel, Basel, Switzerland.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; Y16767; CAA76368.1; -
DR EMBL; Y16766; CAA76367.1; -
DR EMBL; AJ303198; CAC19688.1; -
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 501 AA; 54066 MW; 2BDD6692C39AC4A CRC64;

Query Match 41.2%; Score 1126; DB 10; Length 501;
Best Local Similarity 44.6%; Pred. No. 4.9e-79;
Matches 225; Conservative 102; Mismatches 142; Indels 36; Gaps 8;

QY 18 GAAAVVDHVAIPISLGRLLAGVAGVQYQGWALQLSLTPYVQTLGLSHALTSPFMLCGP 77
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 GATAAPPSPRSVLSRLLRVAVSACIQFGWALQLSLTPYVQTLGLSHALTSPFMLCGP 73
QY 78 TAGLVQPLVGLYSDRCTSRWRRRPFILITGCMILICVAVIVVGFSSDGAALGDTKEHCS 137
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 LSGLLVQPIVGHMSDQCTSKYGRRPFIVAGGTAILAVIIAHSDIGLLGDTADNKT 133
QY 138 LYHGPRHAAIVVVLGFWLLDFSNNTVOGPARAMMADLC--DHHGPSAANSIFCSNMA 195
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 134 M-----AIVAFVIGFWILDVANNMTQGPCRALLADLTGNDARRTRVANAYFSLEFMAIG 186
QY 196 NILGSSGNTNNHKWFFPKTTSACCEACANLKGAFVAVVFLVCLVTTLIFAKEVPYR 255
Db 187 NVLGATGAGVGHYKVPFSLTSCINCANLKSAYIDIIFIITYISISAKERPRI 246
QY 256 ANENLPTTKAGGEVET---EPTGPLAVLKGFKDLPFGMPVSVLLVTAITWLSWFFPFIYDT 312
Db 247 SSQDGPQFSDGTQAQSHIEEAFWLWELFGTFRLLPGSVVILLVTCNNWIGWFFPFIYDT 306
QY 313 DWNGREIYHCDPKGSNAQISAFNEGVRVGAFLGALLNSVILGFSFLIEPMCRKVGRPVW 372
Db 307 DWNGREIYGEPN-----QGOQSYSDGVRMGAFGLMNSVILGITSVLMKLCRWGSGFMW 362
QY 373 VTSNFMVCMATAALISFWSLR-DY-----HGYVQDAITANASIKAVCLVLFAPLGV 425
Db 363 GLSNILMTICFFAMLLITFAKNMDYGTNPPNGVISA-----LIYFALIGIP 411
QY 426 LAILYSVPFVAVTAQLAATRGGGGLCTGVNLISIVIPQVITIALGAGPMDALFGKGNIPAF 485
Db 412 LAITYSPVALVSTRIESLGLQGLSGVNLAIWVQVIVSLVSGSPWDQLFGGNSPAF 471
QY 486 GVSAPALGVGVGVFLL--PKISK 508
Db 472 VVAALSAFAGLIALAIRPRVDK 496

RESULT 16
Q9S725
ID Q9S725 PRELIMINARY; PRT; 512 AA.
AC Q9S725;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE Sucrose transporter SUT2B.
GN SUT2B OR SUT2A.
OS Apium graveolens (Celery).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Apium.
OX NCBI_TaxID=4045;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MATURE LEAF;
RA Noiraud N., Lemoine R.;
RT "Sucrose transporters in celery.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AF167416; AAD45391.1; -
DR EMBL; AF167415; AAD45390.1; -
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; tigr01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 512 AA; 54426 MW; 263753216FFLED2 CRC64;

Query Match 41.1%; Score 1122.5; DB 10; Length 512;
Best Local Similarity 46.3%; Pred. No. 9.5e-79;
Matches 228; Conservative 99; Mismatches 134; Indels 31; Gaps 11;

QY 28 PISGLRILAGVAGGVQVQWALQSLTTPYVOTLGLSHALTSPFMLCGPIAGLVQPLVSDR 87
Db 28 PIATWKLILVAATAAGVQFQWALQSLTTPYVQELGIPAHAWSVIWLCLGSLGLVQPLVGHSSDR 102
QY 94 CTSRWGRRRRPILTCMLICVAVIVVGFSSDIGAALDGTKEHCSLYHGRHAAIVVVLG 153
Db 103 CTSKYGRRRRPFIAGAVAISIVSVVIGHAADIGHAFDREGKIK----PR--AIVAFVLG 156
QY 154 FWLDFSNNTVQGPARAMMADLC--DHHGPSAANSIFCSMMALGNILGYSSGNTNNHKW 211
Db 157 FWILDVANNMTQGPCRALLADLTENDNRTRRVANGYFSLPMAVGNVLGYATGYSNGWYKI 216
QY 212 FPELKTSAACEACANLKGAFVAVVFLVCLVTTLIFAKEVPYRANENLPTTKAGGEVET 271
Db 217 FTETKTACVNECANLKSAYIDIYVFIATITLSVSAHEVPLAS----LTSEAHGOTSG 272
```

```
Db 201 SYNLYKLPFKSTHACDLYCANLKSFIISVLLIEITVLALTVVREKQWSPDEADDEP 260
QY 261 PTTKAGGEVET--EPTGPLAVLKGFKDLPFGMPVSVLLVTAITWLSWFFPFIYDTDNWGRE 318
Db 261 PSS---GKIDVFGLFGAL-----KDLPRMLMLLVVTCNNWIAWFFPFIYDTDNWGRE 311
QY 319 IYHGDPKGSNAQISAFNEGVRVGAFLGALLNSVILGFSFLIEPMCRKV-GPRVVVWTSNF 377
Db 312 IY---GGTAGQKLYDQGVYRGSGLGALLNSVILGTSIAVEYLVRGVGGVKILWGLVNF 367
QY 378 MVCVMAATAALISFWSLRD-YHGYVQDAITANASIKAVCLVLFAPLGVPLAILYSVPFV 436
Db 368 LLAIGLVMTVWVSKVAQHQHGAQGLLPSPSACVKAAGALSLSILGILPLSITFSIPFAL 427
QY 437 TAQLAATRGGGGLCTGVNLISIVIPQVITIALGAGPMDALFGKNIIPAFGVASAFALVGG 496
Db 428 ASIYSSGSGAGQGLSLGVNLAIWVQVIVSLVSGPDSLFGGGLNLPFVVGVAISAISG 487
QY 497 VVGVLPLPKISK 508
Db 488 VLAIVLLPKPSK 499

RESULT 17
Q9FE59
ID Q9FE59 PRELIMINARY; PRT; 510 AA.
AC Q9FE59;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Sucrose transporter SUT4.
GN SUT4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsals.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG ERECTA;
RX MEDLINE=20407148; PubMed=10948254;
RA Weise A., Barker L., Kuhn C., Lalonde S., Buschmann H., Frommer W.B.,
RA Ward J.M.;
RT "A new subfamily of sucrose transporters, SUT4, with low affinity/high
RT capacity localized in enucleate sieve elements of plants.";
RL Plant Cell 12:1345-1355(2000).
DR EMBL; AF175322; AAG09192.1; -
DR EMBL; AF175321; AAG09191.1; -
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; tigr01301; GPH_sucrose; 1.
SQ SEQUENCE 510 AA; 54835 MW; 986C68016D4CAEA4 CRC64;

Query Match 41.1%; Score 1122; DB 10; Length 510;
Best Local Similarity 47.6%; Pred. No. 1e-78;
Matches 226; Conservative 85; Mismatches 144; Indels 20; Gaps 8;

QY 34 LILAGVAGGVQVQWALQSLTTPYVOTLGLSHALTSPFMLCGPIAGLVQPLVGLYSDR 93
Db 43 LLRVASVACGQFQWALQSLTTPYVQELGIPAHAWSVIWLCLGSLGLVQPLVGHSSDR 102
QY 94 CTSRWGRRRRPILTCMLICVAVIVVGFSSDIGAALDGTKEHCSLYHGRHAAIVVVLG 153
Db 103 CTSKYGRRRRPFIAGAVAISIVSVVIGHAADIGHAFDREGKIK----PR--AIVAFVLG 156
QY 154 FWLDFSNNTVQGPARAMMADLC--DHHGPSAANSIFCSMMALGNILGYSSGNTNNHKW 211
Db 157 FWILDVANNMTQGPCRALLADLTENDNRTRRVANGYFSLPMAVGNVLGYATGYSNGWYKI 216
QY 212 FPELKTSAACEACANLKGAFVAVVFLVCLVTTLIFAKEVPYRANENLPTTKAGGEVET 271
Db 217 FTETKTACVNECANLKSAYIDIYVFIATITLSVSAHEVPLAS----LTSEAHGOTSG 272
```

```
Qy 272 EPTGFLAVLKG-FKDLPPGMPSVLIVTAITWLSWPFILYDWDGRIYHGDPRGKSNQ 330
Db 273 TDEAFLEIFGFRFYFPGVNVIIILVLTALTWIGWPFILFDWDGRIYGEIPNIG--- 329
Qy 331 ISAFNEGVRVGAFLGLLSVILGRSSFLIEPMCKRKVGRVNVVTSNFMVCMVAMATALIS 390
Db 330 -TSVAGVSMGALGMLNSVFLGITSVLMEKLCRKWAGFGWGISNIIILMAICFLGMIITS 388
Qy 391 FWSLRDHYGVQDITANASTKAVCLVFLAFGLVPLATLYSVPAVTAQLAATRGGOGL 450
Db 389 F--VASHLGYIGHE-QPPASIVFAVLFTITGLIPLATYSPVALISIRIESLGLGOGL 445
Qy 451 CTGVLNISIVPOVIALGAGPMDALFGKGNIPAFGVASAFALVGVGVVFLPK 505
Db 446 SLGVNLALVIVQVIVSGVSGPMDOLFSGGNSPALAVGAATGFGIVAILALPR 500

RESULT 18
Q8VYX4 PRELIMINARY; PRT; 513 AA.
AC Q8VYX4;
DT 01-MAR-2002 (TrEMBLrel. 20, created)
DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE sucrose transporter SUC1.
GN SUC1.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RA Coupe S.A., Sinclair B.K., Bucknell T.T., Greer L.A., Eason J.R.,
RA Heyes J.A.;
RT "The isolation and characterization of sucrose transporter homologs
RT from broccoli and their role in sucrose mobilization during
RT senescence.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY065839; AAL58071.1; -
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
SQ SEQUENCE 513 AA; 54663 MW; 0279B262291B2B CRC64;

Query Match 41.18; Score 1122; DB 10; Length 513;
Best Local Similarity 43.78; Pred. No. 1e-78;
Matches 220; Conservative 103; Mismatches 156; Indels 24; Gaps 6;

Qy 10 AELSAGVGAADVHDVAPISLGRILLAGMVAGGVQYQWALQSLTTPYVQTLGLSHALT 69.
Db 10 AKDAAALETQSSLEEFNQPSPLRKTIIVSSIAAGVQFGWALQSLTTPYVQTLGLPHKWS 69
Qy 70 SFMWLCGPIAGLVQPLVGLSDRCTSRWRRRPFLTGCMILICVAVIVGFGSSDICAL 129
Db 70 SLIWLICGVSMIVQPIVGFHSDRCTSRFRRRPFIAGAAMVAVFLGIYAADIGYKM 129
Qy 130 GDTKEHCSLYHGRWHAIVVGLFWLLDFSNNTVQGPARAMMADLC--DHHGPSAANSI 187
Db 130 GDKLEQT-----PRVALGIFALGVLDVANNTLQGPCRAFLADLAAGDAKRTRVANAF 184
Qy 188 FCSWMALGNILGYSSTNNHKKVPELTKTSACCEACANLKGAFLVAVVFLVLCITVTLI 247
Db 185 FSFFMAVGNVLGYAGSTFNLHKMPFPFAMTACDLYCANLKSCTFLSITLLIIVTVSLW 244
Qy 248 FAKE-----VYRANENLPTTKAGEVETPTGLAVLKGDKLPDPPGMPSVLLVTAITWL 302
Db 245 YVKDKQWSPPPVAADEEKKSPFFGE-----IFGAFKVMRPMWMLLIIVTALNWI 294
Qy 303 SWFEPILYDWDGRIYHGDPRGKSNQISAFNEGVRVGAFLGLLSVILGRSSFLIEPM 362
Db 295 AWFPELFDWDGRIYHGDGSDARLQYINKGVQSGALGMLNSVILGFNSLGVSEWI 354
```

```
Qy 363 CRKV-GPRVWVVTNFMVCMVAMATALISFWSLRDHYGVQDITANASTKAVCLVFLAF 421
Db 355 GKVGKARLWGIWNFIIAIGLAMTVLVTKLA-ADYRKVAGPYAGSPGIRAGALSIFAV 413
Qy 422 LGVPLATLYSVPAVTAQLAATRGGOGLCTGVLNISIVIPQVIALGAGPMDALFGKGN 481
Db 414 LGIPLATFESIPFALASIFSSGAGOGSLGVNLALVIVQVIVPMIVSLVGGGPDALFGGN 473
Qy 482 IPAFGVASAFALVGVGVVFLLP 504
Db 474 LPAFIVGAIAAISGVLAITVLP 496

RESULT 19
Q9ZTB9 PRELIMINARY; PRT; 512 AA.
AC Q9ZTB9;
DT 01-MAY-1999 (TrEMBLrel. 10, created)
DT 01-MAY-1999 (TrEMBLrel. 10, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE Sucrose transporter.
GN SUT1.
OS Apium graveolens (Celery).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Apium.
OX NCBI_TaxID=4045;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-LEAF;
RA Noiraud N., Delrot S., Lemoine R.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF063400; AAC99332.1; -
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRFAMS: TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 512 AA; 54520 MW; 4D76A7854A7DF903 CRC64;

Query Match 41.08; Score 1120.5; DB 10; Length 512;
Best Local Similarity 45.48; Pred. No. 1.3e-78;
Matches 222; Conservative 100; Mismatches 142; Indels 25; Gaps 8;

Qy 28 PISLGRILLAGMVAGGVQYQWALQSLTTPYVQTLGLSHALTSEMWLCGPIAGLVQPLV 87
Db 28 PIATWKLILVAAIAAGVQFGWALQSLTTPYVQTLGLPHKWAAYIWLGGPISGMLVQPIV 87
Qy 88 GLYSDRCTSRWRRRPFLTGCMILICVAVIVGFGSSDICALGDTKEHCSLYHGRWHA 147
Db 88 GYSDRCQSSFRRRPFIASGAGCVATSVILIGFAADIGYKAGDDMNKTL---KPR--AV 142
Qy 148 IVYVLGFWLLDFSNNTVQGPARAMMADLC--DHHGPSAANSIFCSWMALGNILGYSST 205
Db 143 TVFVIGFWLDVANNTLQGPCRALLDLCNGDTRMRSANALYRFFMAVGNILGNAAGSY 202
Qy 206 NNHWKWPFLKTSACCEACANLKGAFLVAVVFLVLCITVTLIFAKEVYRANENLPTTKA 265
Db 203 NNLYKLFPFSTHACDLYCANLKSCTFLISVILLITVLTAVVREKOWSPDE----- 255
Qy 266 GGEVETPTGLAV----LKGFDLPPGMPSVLIVTAITWLSWPFILYDWDGRIYH 321
Db 256 -ADEEPPSSGKIPVFGELLRALKDLPRPMLMLLAVTCLNWTIAWFPFIFLFDWDGRIY- 313
Qy 322 GDPKGSNAQISAFNEGVRVGAFLGLLSVILGRSSFLIEPMCKRV-GPRVWVVTNFMV 380
Db 314 ---GGTAGQGLYDQGVVGVSLGLLSVILGLTSIAVEYLVVRGVGVKILWGLVNFLLA 370
Qy 381 VAMATALISFWSLRD-YHGVQDITANASTKAVCLVFLAFGLVPLATLYSVPAVTAQ 439
Db 371 IGLVWTVVSKVAQHQRHANGQLPPSAGVKGAGALSFLSILGIPLSITSIFFALASI 430
Qy 440 LAATRGGOGLCTGVLNISIVIPQVIALGAGPMDALFGKGNIPAFGVASAFALVGVGV 499
```

```

Db 431 YSSGAGAGLGLSLVNLAIWVPMIVSLVGLPPDLSLFGGNLPFAVYVYGAISAISGVLA 490
Qy 500 VELLPKISK 508
Db 491 IVLLPKPCK 499

RESULT 20
Q9M3R4 PRELIMINARY; PRT; 510 AA.
AC Q9M3R4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Sucrose transporter (Putative sucrose/H+ symporter).
GN SUC4 OR AT1G09960.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
RN NCBI_TaxID=3702;
RP SEQUENCE FROM N.A.
RA Sauer N.K.;
RT "An Arabidopsis thaliana sucrose transporter.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2].
RP SEQUENCE FROM N.A.
RA Yamada K., Banh J., Chan M.M., Chang C.H., Chang E., Dale J.M.,
RA Deng J.M., Goldsmith A.D., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Wu H.C., Yamamura Y., Yu G., Bowser L.,
RA Karlin P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Shinozaki K.,
RA Davis R.W., Ecker J.R., Theologis A.;
RA "Full Length cDNA of gene At1G09960 (GI:15218362).";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ289166; CAB92308.1;
DR EMBL: AY072092; AAL59915.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRfams: TIGR01301; GPH_sucrose; 1.
SQ SEQUENCE 510 AA; 54784 MW; 3AF99CF2D48ACBD4 CRC64;

Query Match 40.9%; Score 1117; DB 10; Length 510;
Best Local Similarity 47.4%; Pred. No. 2.5e-78;
Matches 225; Conservative 85; Mismatches 145; Indels 20; Gaps 8;

Qy 34 LILAGVAGGVQYQWALQSLTPYVQTLGLSHALTSFMMLCGPAGLVVQPLVGLYSDR 93
Db 43 LLRVASVACGIFQWALQSLTPYVQTLGLSHALTSFMMLCGPAGLVVQPLVGLYSDR 102
Qy 94 CTSRWGRRRPFILTCMLICVAVIVGVFSSDGAALGDTKEHCSLYHGPRHAAIVVVLG 153
Db 103 CTSYGRRRRPFIVAGAVAISVNVIGHAADIAGWAFDREGKIK----PR--AIVAEVLG 156
Qy 154 FWLLDFSNNTVQGPARAMMADLC--DHHGPSAANSIFCSHMLGNILYSGSGSTNNHKK 211
Db 157 FWLLDVANNMTQGPCRALLADLTENDNRRTRVANGYSLEFMAVGNVLGYATGNGWYKI 216
Qy 212 FPELKTSAACEACANLKGAFVAVVFLVCLTFLIPAKEVPYRANENLPTTRAGGVEVET 271
Db 217 FTFKTVACNVCANLKSAYIDVFTAITILSVSAHEVPLAS-----LASEAGGTSG 272
Qy 272 EPTGPLAVLKG-FKDLPPGMPSSLVTAITWLSWFFPILYDTDMGMEIYHGDPKGSNAQ 330
Db 273 TDEAFLSEIFGTRYFPNGWIIILLVTAIWIGWFFPILFDTDMGMEIYGGEPNIG--- 329
Qy 331 ISAFNEGVVRVAGCLLNSVLGFSFSLIEPMCRKVKVPRVWVTSNFMVCMVMAATALIS 390
Db 330 -TSYSGVSMGALGLMLNSVFLGITSVLMKLCRKWAGFVWGISNLMALICFLGMIITS 388

```

```

Qy 391 FWSLRDHYGVQDDAITANASTKAVCLVLEFAPLVGLVPLATLYSPFAVTAQLAATRGGOGL 450
Db 389 F--VASHLGTIGHE-QPPASIVFAVLTITTLTGLPILAITISVPALISIRIESLUGGOGL 445
Qy 451 CTGVLNISVIPIQVITIALGAGPMDALFCKGNIPAFGVASAFALVGVGVVFLPK 505
Db 446 SLGVLNLAIPIQVIVSGSGPMDQLFEGGNSPALAVGAATFGIGIVAILALPR 500

RESULT 21
Q40938 PRELIMINARY; PRT; 510 AA.
AC Q40938;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Sucrose transporter.
GN PTP1.
OS Plantago major (common plantain).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Veroniceae; Plantago.
RN NCBI_TaxID=29818;
RP SEQUENCE FROM N.A.
RA MEDLINE=95093474; PubMed=8000426;
RA Gahriz M., Stolz J., Sauer N.;
RT "A phloem-specific sucrose-H+ symporter from plantago major L.
RT supports the model of apoplastic phloem loading.";
RL Plant J. 6:697-706(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: X75764; CAA53390.1;
DR InterPro: IPR003662; sub_transporter.
DR Pfam: PF00083; sugar_tr; 1.
DR TIGRfams: TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 510 AA; 54018 MW; C6905E27F9A80B85 CRC64;

Query Match 40.9%; Score 1116; DB 10; Length 510;
Best Local Similarity 44.0%; Pred. No. 3e-78;
Matches 227; Conservative 103; Mismatches 158; Indels 28; Gaps 11;

Qy 18 GAAAVV-----DHVAPISLRLILAGVAGGVQYQWALQSLTPYVQTLGLSHALTSF 71
Db 9 GAKTIITTPPEEAPIR--NIFLVAIAAGVQFGWALQSLTPYVQTLGLSHALTSF 66
Qy 72 MWLCGPTAGLVVQPLVGLYSDRCSRWGRRRPFILTCMLICVAVIVGVFSSDGAALGD 131
Db 67 IWLCPISGIMVQPVGVFSDNCTSFREGRRRPFITAGAGLVGVAVVLIGFAADLGHAGGD 126
Qy 132 TKHCSLYHGPRHAAIVVVLGFWLLDFSNNTVQGPARAMMADLCDDHGP--SAANSTFC 189
Db 127 -----SLGDGLKPRAGVVFVFGFWLIDVANNMLOGPCRALLDLGGTKKMANANSFFS 181
Qy 190 SMWALGNILYSGSGSTNNHKKWFFPKTSACCEACANLKGAFVAVVFLVCLTFLVIFA 249
Db 182 FEMAVGNVLGYAAGSVSMYKVPFESKTKACDIYCANLKSFIITLITLITLTLALSIV 241
Qy 250 KEVPIRANENLPTTKAGEVETEPTGLAVLKGFDLPFGMPSSLVLTATWLSWFFPIL 309
Db 242 REKRHVAEEQVTAARKGFKIPVFP----ELFGALKDLPMPMWLVLLVTALNIAWFGFL 297
Qy 310 YDTDMGREIYHGDPKGSNAQISA--FNEGVRVCAFGALLNSVLGSSSFLIEPMCRKV- 366
Db 298 FDTDMGREYV-GETQHKAPELAVIYKNGVSAGALGLMLNSIVLGFASLGVQMARALG 356
Qy 367 GPRVWVTSNFMVCMVMAATALISFWSLRDHYGVVQDAI--TANASIKAVCLVLEFV 424
Db 357 GVRKLGWVNFILAIICLMTIVIT--KVASHHRPYSNCVLOTPESSVKIGALVVSALGI 414
Qy 425 PLATLYSVFPAVTAQLAATRGGOGLCTGVNLINISIVIPQVITIALGAGPMDALFKGNIPA 484

```

Db 415 PLATFSPVPALASIYSTTTGSCQGLSLGVNLNLAIVIPQMIVSVASGPWDAMFGGNLPA 478

Qy 485 FGVAFAEALYGGVVGVELLPKIKSKROFRAYS-AGGH 519
I : I : I :
I : I : I :

Db 475 FVUGAVAAAAGSIFAFTMLPSPAESIKNLSVAGGH 510

RESULT 22

QX9HL6 PRELIMINARY; PRT; 524 AA.

AC AC9XHL6;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Sucrose transport protein SUT1.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxId=3888;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99291546; PubMed=10363367;
RA Tegeder M., Wang X.D., Frommer W.B., Offler C.E., Patrick J.W.;
RT "Sucrose transport into developing seeds of *Pisum sativum* L.";
RC Plant J. 18:151-161(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY);
DR EMBL; AF109922; AAD41024.1;
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR TIGRfams; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 524 AA; 55279 MW; B5CE3F880D2C1E6B CRC64;

Query Match 40.9%; Score 1116; DB 10; Length 524;
Best Local Similarity 44.2%; Pred.No.3.1e-78;
Matches 221; Conservative 106; Mismatches 153; Indels 20; Gaps

Qy 28 PISLRLLLAGWAGVGYGWALQLSLTTPYVOTGLSHALTSMFMLCGPFIAGLVVQPIV 87
I : : : : :
I : : : : :
Db 32 PSPLRKTMVVASIAAGVQFGMALQLSLTTPYVOLLGHHHTWAAYIWLCGPISGMVQPVV 91
I : : : : :
I : : : : :
Qy 88 GLYSRCTSRWGRRRPILTGCMLICVAVTVWGSSDIGAAGDYTEHCSLYHGPKRWHA 147
I : : : : :
I : : : : :
Db 92 GYHSRCTSRFGRRRPPIAAGSAVAIAVLPGYAADLGHSFGD----NLDKKVRPRI 146
I : : : : :
I : : : : :
Qy 148 IVYVLGFWLDFSNNTVOGPARAMMADIC--DHGSPSAANSIFCSMWALGNLTGYSSGST 205
I : : : : :
I : : : : :
Db 147 GIFVVGWEILDVANMMIQGPCRALLDJLCAGNRKTNNANAFSFPMGVNVLGYAAGY 206
I : : : : :
I : : : : :
Qy 206 NNHWKHPFPFKTSACCACANLKAGFLVAVVFVLCITVTLIFAKEPVPRANENLPTTKA 265
I : : : : :
I : : : : :
Db 207 SKLYHVFPFKTEACNVYCANLKSCEFSLIALTLVLTAAIIYVKETPLIAEKAVVTAED 266
I : : : : :
I : : : : :
Qy 266 GGEVETEPT-GPIAVLKGFDLPFGMPSSLVLLTAITWLSWFFILYDTDMMGREIYHGD 324
I : : : : :
I : : : : :
Db 267 GSGNGMPCFCQLS--GAPKELRKPWWILLVTCNLNWIAWFPELLEDDTDMGKEYV--- 320
I : : : : :
I : : : : :
Qy 325 KGSNAQTASRNEGVRVCAFGLLNSVILGFSSFLIEPMCCKV-GPRVVVWTFNFVCVM 383
I : : : : :
I : : : : :
Db 321 GGTVGEGHAYDMGVRAGALGMLNSVVLGATSLGVSDILARGVGKRLMGIVNFLAI 380
I : : : : :
I : : : : :
Qy 384 AATAALISFWS--LRDY--HGYYODAITNASIKAVCLVLFALFLGVPLAILYVSPFAVTA 438
I : : : : :
I : : : : :
Db 381 GLTVLVTKLAQHRSQRVAPGTGGQQDPPLPSGGIKAGALLTFSVLGIPLAITYSIPPALAS 440
I : : : : :
I : : : : :
Qy 439 QLAATRGCGGLCTGVLNISIVIPQIIALFGKWPDALFCGNIPAFGVASAFALYGGVV 498
I : : : : :
I : : : : :
Db 441 IFSSTSGAGQGLSIGVNLNLAIVIPQMFVSVLSPWDALFGGNLPAFVVGAVAALASGIL 500
I : : : : :
I : : : : :
Qy 499 GVFLLPKISKRQFRAYSAGG 518
I : : : : :
I : : : : :
Db 501 SMILLPPSPPDMAKSVSATG 520
I : : : : :
I : : : : :

[illegible]

RESULT 25	
065803	
ID	065803
AC	065803
DT	01-APR-2010
DT	01-APR-2010

